LLNL-JRNL-416534

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Superresolution with Seismic Arrays using Empirical Matched Field Processing

D. B. Harris, T. Kvaerna

September 3, 2009

Geophysical Journal International

Disclaimer

Superresolution with Seismic Arrays using Empirical Matched Field Processing

David B. Harris[1], Tormod Kvaerna[2]

[1]Lawrence Livermore National Laboratory, [2]NORSAR

Accepted: <*date*> Received: <*date*>; in original form <*date*>

Suggested abbreviated title: Supperresolution with Matched Field Processing

Corresponding author: David B. Harris
P. O. Box 231
Maupin, 0R 97037
USA
1-510-331-8804
oregondsp@gmail.com

Summary

Scattering and refraction of seismic waves can be exploited with empirical matched field processing of array observations to distinguish sources separated by much less than the classical resolution limit. To describe this effect, we use the term "superresolution", a term widely used in the optics and signal processing literature to denote systems that break the diffraction limit. We illustrate superresolution with Pn signals recorded by the ARCES array in northern Norway, using them to identify the origins with 98.2% accuracy of 549 explosions conducted by closely-spaced mines in northwest Russia. The mines are observed at 340 - 410 kilometers range and are separated by as little as 3 kilometers. When viewed from ARCES many are separated by just tenths of a degree in azimuth. This classification performance results from an adaptation to transient seismic signals of techniques developed in underwater acoustics for localization of continuous sound sources. Matched field processing is a potential competitor to frequency-wavenumber and waveform correlation methods currently used for event detection, classification and location. It operates by capturing the spatial structure of wavefields incident from a particular source in a series of narrow frequency bands. In the rich seismic scattering environment, closely-spaced sources far from the observing array nonetheless produce distinct wavefield amplitude and phase patterns across the small array aperture. With observations of repeating events, these patterns can be calibrated over a wide band of frequencies (e.g. 2.5 – 12.5 Hertz) for use in a power estimation technique similar to frequency-wavenumber analysis. The calibrations enable coherent processing at high frequencies at which wavefields normally are considered incoherent under a plane wave model.

Introduction

Permanent seismic arrays are part of the verification regime for nuclear test ban treaties, especially the Comprehensive Test Ban Treaty (CTBT). Interest in smaller explosions is pushing the monitoring threshold to increasingly lower levels. Since the number of events grows exponentially with decreases in magnitude, reliable methods of screening large numbers of small events are required if analysis resources are not to be overwhelmed. With the capacity to estimate the direction and speed of observed waves and to detect small signals, arrays are essential assets in screening smaller events, especially explosions, by attributing them to known sources.

However, the performance of seismic arrays in resolving sources is fundamentally limited by signal processing algorithms that employ plane-wave assumptions. The propagation environment is so strongly heterogeneous, and scattering so pronounced, that the wavefields of regional events can be modeled as planar only over apertures of one to two wavelengths, constraining array geometries and limiting resolution (Mykkeltveit et al., 1983). The performance of, specifically, frequency-wavenumber (FK) analysis (Capon, 1969), and beamforming (Johnson and Dudgeon, 1993) are limited by scattering effects. FK analysis decomposes the incident wavefield in the frequency domain as a superposition of complex exponential plane waves and provides a map (spectrum) of power or energy incident on the array as a function of wavenumber. It is the principal tool for ascertaining the directions and velocities of waves incident on a seismic array. These parameters are obtained by extracting the vector wavenumber corresponding to peaks in the FK spectrum. Since seismic arrays operate in an inhomogeneous propagation environment, peaks in the FK spectrum often are not well defined. The problem is acute at higher frequencies.

Work on correlation detection and classification using array and network waveforms indicates that problems associated with scattering can be overcome by calibration of signal characteristics at the array aperture (Gibbons and Ringdal, 2006; Gibbons, et al., 2007; Harris, 1991). Waveform correlation methods obviate loss of coherence across an aperture by calibrating simultaneously the temporal and spatial structure of signals from specific repeating sources. Discoveries that large subpopulations of explosions and earthquakes produce nearly identical repeating waveforms (Israelsson, 1990; Schaff and Richards, 2004; Schaff and Waldhauser, 2005) motivate large-scale application of waveform correlation techniques for high precision location (see e.g. Waldhauser and Ellsworth, 2000; Hauksson and Shearer, 2005), classification, and detection.

However, the applicability of correlation methods is limited by variations in source characteristics (time history, mechanism and physical dimension) which change the observed waveform. This problem motivates a search for invariants in the array waveform for a particular source, unaffected by source variations. For sources that vary principally through changes in time history, the collection of Greens functions that characterize propagation from the source to the array stations are the relevant invariant. At the array aperture, and in narrow frequency bands, collections of Greens functions determine the spatial structure of the signal.

In the underwater sound community, matched field processing (Bucker, 1976; Baggeroer et al., 1993) extends coherent array processing methods to more heterogeneous environments by characterizing the spatial structure of the signal under propagation conditions more general than the free-space assumption that leads to plane-wave structure. The essence of matched field processing (MFP), as practiced in underwater sound, is to calculate the Greens functions describing propagation from source to array receivers using an accurate model of sound velocity in the oceanic waveguide. MFP steers the array with the more accurate (non-planar) signal representation that results to focus incident signal energy emanating from a particular source point in a manner directly analogous to FK analysis and beamforming. Recall that beamforming methods steer an array by applying time delays to align waveforms recorded by the array elements. The time delays are projections of the slowness vector of a presumed incident plane wave onto the array sensor positions. In the frequency domain this operation corresponds to multiplying the Fourier representations of the sensor waveforms by corresponding complex phase factors. MFP, by contrast, steers the array explicitly in the frequency domain using the complex phase and amplitude factors obtained by solving the wave equation through the propagation model. Because the sources of interest in underwater sound investigations usually are ships with rotating machinery, the signals often are narrowband or are comprised of a series of narrowband components. Consequently, Greens functions typically are calculated for a single frequency by solving the Helmholtz equation (Baggeroer et al., 1993).

MFP offers the prospect of much better detection and estimation performance than beamforming or FK analysis because the representation of the signal can be much more realistic. In the underwater sound problem, superior performance often is achieved by matched field methods, especially in deep water applications. However, MFP has not found significant application in seismology due to the difficulty of developing realistic earth models to predict the structure of seismic wave fields at frequencies much above a tenth of a Hertz.

An alternative to calculating the wavefield structure across an array is to estimate that structure directly from field calibration data (Fialkowski et al., 2000). We refer to this strategy as empirical matched field processing. Attempts at empirical MFP in underwater acoustics have achieved mixed success due to the dynamic nature of the medium and underwater sound sources. The principal issue is that stable estimates of the vector of amplitude-phase steering factors (i.e. the matching field, also called the steering vector) are difficult to obtain for non-stationary sources. In the classic MFP approach, the steering vector is obtained as the principal eigenvector of a sample covariance matrix estimated from calibration data over the collection of array sensors. Stationarity is assumed, and using ergodic properties, the sample covariance is averaged over a long time window to achieve statistical stability. Typically, the continuous data are segmented into a series of contiguous "snapshots", a sample covariance matrix for the array is estimated for each snapshot and the final estimate is the average of sample covariance matrices over all snapshots. However, in the empirical MFP application, the calibration source moves to cover a source region of interest. This motion limits the source dwell time in a particular location and the corresponding covariance integration time.

Of course, the seismic medium is fixed as are many seismic sources, such as mines. This fact provides the opportunity for empirical MFP to succeed in seismic applications by averaging covariance matrix estimates over ensembles of events in a particular source region.

In this paper we describe an adaptation of empirical MFP to operate on transient seismic signals. The application we present in detail most closely resembles FK analysis. Consequently, we will test MFP and contrast its performance to that of FK analysis in the context of classifying events by their origins at a set of discrete locations. Our principal result is that MFP performs much better as an event classifier than FK methods, even than FK methods with the sort of empirical corrections for refraction that are improving array operations (Schweitzer, 2001). Indeed, we show that we can reliably distinguish events from mines that are separated by much less than the classical (Rayleigh) resolution limit as viewed from a distant, regional array. This performance is obtained without the sensitivity to source time history variations that would defeat waveform correlation classifiers.

The principal hurdle to be overcome in adapting MFP to the seismic application is that matched field processing, as practiced in underwater sound, naturally applies to long observations of monochromatic sources. The seismic signal, on the other hand, typically is relatively wideband and transient. In our adaptation, the method uses repeating events at a known source to calibrate the complex *spatial* structure of the wavefield incident upon the array in a series of narrow frequency bands. Decomposing the array signal into narrow frequency bands effectively separates spatial and temporal structures of the signal allowing independent exploitation of the spatial structure.

The seismic waveform from a discrete source is a heavily multipathed signal with many phase arrivals representing distinct branches of propagation. The seismic signal is profoundly non-stationary, as each phase has a distinct propagation path and scattering environment. Consequently, the spatial structure of the seismic waveform changes rapidly as phases come and go throughout the seismogram. To simplify analysis in this first application of MFP to seismic data, we restrict our analysis to a single branch of propagation, the first-arriving Pn phase of regional observations. However, in the conclusions, we will comment upon a general structure for MFP that extends to the entire waveform, which we expect will provide opportunities in detection.

Since the Pn temporal window is too short to permit significant snapshot averaging with a single event observation, we obtain stable covariance matrix estimates by averaging the covariance over event ensembles,. This approach leads us to reconsider the stochastic representation of seismic observations commonly assumed in FK analysis. The FK spectrum, as classically defined, describes the power density in frequency-wavenumber space of a process which is a temporally stationary and spatially homogeneous random field (Capon, 1969). However, over time FK analysis has come to be applied to signals from individual events with very short temporal windows. Consequently, the FK spectrum as commonly used, is the energy density spectrum (the squared magnitude of a Fourier transform) of a transient, implicitly deterministic signal. In this paper, we introduce a probability model for the seismic waveform which is temporally non-stationary and spatial inhomogeneous, resulting from the convolution of transient random forces in the source region with deterministic, but unknown Greens functions.

As a concrete example, we study the European arctic region (Figure 1) where the ARCES array, an important CTBT monitoring station, is 340-410 kilometers from two groups of mines in the Kola peninsula, Russia: the Olenegorsk (O1-O5) and Khibiny (K1-K5) groups. In monitoring for signals from distant nuclear events, the array observes thousands of mining explosions annually, which can be screened if attributed to their originating mines. However, the array, with a 3 kilometer aperture, is too small to resolve many individual mines grouped within some mining regions. Array resolution (by the Rayleigh criterion) is determined by the separation of half-power (3 dB) points of the main lobe of the array wavenumber response. In Figure 1, the 3 dB points of the array response at 4, 8 and 12 Hz for Pn waves are projected onto the arrays geographic field of view. The array has been steered to one particular mine (K2). All ten mines fall within the 3 dB contours even at 12 Hz. Consequently, the array should not, and we will show, does not, reliably distinguish events among these mines using plane-wave methods applied to Pn observations. However, we show that matched field processing methods do distinguish these events with a high degree of reliability. This performance is an indication that the rich scattering environment imprints characteristic spatial structure on the signals that may be captured and exploited empirically.

The paper is organized in three additional sections. We devote the first to mathematical background describing our adaptation of matched field methods to process transient seismic signals. The second section describes the data and analytical methods we use to test and contrast the performance of FK and MFP methods. We pay particular attention to the spatial characteristics of matched field calibrations and show how these differ from the plane wave model used in FK algorithms. The matched field calibrations show interesting structure imprinted by propagation in a strongly heterogeneous medium. The structure is responsible for the ability of matched field methods to distinguish closely-spaced sources at considerable range. The short third section presents classification results. The last section discusses the implications of these results for extending coherent seismic array processing techniques to larger apertures and higher frequencies, and briefly describes methods for generalizing MFP to use the entire waveform.

Mathematical Background

In this study our objective is to identify the source of an event from waveforms observed at an array. We define □ to index the source among a set Θ of possible sources, in our case the 10 mines of the Kola peninsula.

Nomenclature for the observed signals

We describe the wavefield incident upon an array aperture as □□,□=□□□,□+□□,□, where the vector[1] □ indicates the position of a sensor observing the wavefield and □ denotes time. The function □□ denotes the signal from an event at source □ and □ denotes additive noise, considered to be stationary and independent from the signal. For simplicity we are considering a scalar wavefield (for example the vertical component of motion across the array aperture). Extension to vector (three-component) wavefields is straightforward. Since we are considering arrays, the incident wavefield is measured on an aperture sampled at □□ discrete locations □□ ;□=1,⋯,□□. It is convenient to collect the observed waveforms into a vector:

□□=□□1,□□□2,□ ⁝ □□□□,□= □□□+□□=
□□□1,□□□□2,□ ⁝ □□□□□,□+□□1,□□□2,□ ⁝ □□□□,□ (1)

For purposes of exposition, we make the assumption that each of the sources, □, is a point source, with a single repeating mechanism. Under this assumption, □□, can be represented by the simple convolution integral [see, for example, Aki and Richards, 1980]

[1] Note to the editor: because we could not coax Word into implementing GRL standards for matrices and vectors, we are using italics to represent scalars, bold lower-case characters to represent vectors and bold upper case characters to represent matrices.

[illegible] = [illegible] (2)

where [illegible] is the vector of Greens functions describing propagation from the source to the observing sensors and the function [illegible] represents the forcing function of the source, which we also refer to as the source time history. For more complex distributed sources or sources exhibiting multiple mechanisms an approximation involving multiple point sources is possible. We comment upon the complexities that assumption entails in the conclusions.

Matched field processing in underwater sound

In underwater sound applications [Baggeroer, 1993] it is common to use a complex analytic signal representation [see e.g. Franks, 1981] and approximate the forcing function as a slowly-varying complex envelope [illegible] modulated by a complex exponential:

[illegible] = [illegible] (3)

In fact, the forcing function is the real part of the complex analytic function. For simplicity of notation, we will work with the complex analytic representation of all signals throughout the discussion and treat the extraction of real components as implied. The forcing function is considered to be stationary over the observation interval, which means that its statistics (moments) are constant. Provided the envelope function is approximately constant over the duration of the Greens functions, the observed array signal has the form:

[illegible] ≈ [illegible] (4)

where ▯▯▯ is the vector of Fourier transforms of the Greens functions. Matched field processing performs a type of beamforming operation by steering the array with a vector ▯ of complex weights:

▯▯ = ▯▯▯▯ (5)

Here the superscript ▯ denotes complex transpose. In the absence of noise, ▯▯ = ▯▯▯▯▯ and the average power of the resulting beam is

▯▯*2*= ▯▯▯▯▯▯*2*▯▯*2.* (6)

▯ is called the steering vector and usually is normalized to have unit length, i.e. ▯▯▯***=1.*** The average power is maximized when the steering vector is proportional to ▯▯.

In the usual localization problem, source location is estimated by maximizing the average power of the beam over a range of possible origins ▯:

max▯ ▯▯▯*2* = max▯▯▯▯▯▯*2* = max▯ ▯ ▯▯▯ ▯▯

(7)

▯**=** ▯▯ ▯ ▯▯ ▯▯**=** ▯▯▯▯▯▯▯▯▯▯▯▯▯▯*12*

The major components of the process are (1) to compute the covariance matrix ▯ of the received signals, (2) to compute steering vectors ▯▯ as normalized Greens function vectors over a range of possible source locations, and (3) to estimate the source location by maximizing the quadratic form (power) ▯ ▯▯▯ ▯▯ over source location ▯.

The covariance matrix ▯ is the sum of the covariance matrix ▯▯ of the signal ▯▯ and the covariance matrix ▯ of the noise ▯, due to the assumed independence of signal and noise. In addition, ▯▯ is stationary (implying ▯ is constant) due to the assumptions of stationarity for ▯▯

and □□. With the assumption of stationarity, ergodic properties are invoked to estimate □. In the underwater sound problem, the available continuous stream of data is broken into (ideally many) segments, a sample covariance matrix is estimated from each segment, and the resulting sample matrices are averaged to provide a stable estimate of □. The steering vectors usually are obtained numerically by solving the Helmholtz equation for all the possible source locations to the observing array through a model of the propagation medium.

We seek to adapt this processing scheme to locate events with a seismic array. Two factors bar simple emulation of the scheme. The first is that models of the seismic propagation environment are inadequate for estimating the Greens functions, except at very low frequencies far below the band where we will demonstrate processing gain. The second is that the seismic source, □□, though possibly still treatable as a random process, has a duration that is very short compared to the duration of the Greens functions observed at regional distances. This is especially true for the smaller events that are of increasing importance in current monitoring applications. Consequently the observations of a seismic signal are strongly non-stationary. Arguments based on ergodic notions cannot be invoked to obtain stable estimates of covariance matrices. We discuss these two factors in turn.

As mentioned in the introduction, one remedy for the problem of inadequate models is to obtain estimates of steering vectors empirically, i.e. from field calibration data [Fialkowski et al., 2000]. In the underwater sound problem, this approach is motivated by the structure of the ideal covariance matrix □≈□ □ when noise is absent and when the array is observing a signal emanating from a known source □. From equation 4:

□□ ≈ □□2□□□□□□□□□□ (8)

we see that the covariance matrix has a single eigenvalue and corresponding eigenvector proportional to □□. Consequently, a suitable steering vector can be obtained as the principal eigenvector of the covariance matrix estimated from ground-truth training data.

However, in the seismic context, the second factor – the short source time history – prevents estimation of signal covariance by averaging over time. Our solution to this problem involves substituting averages over ensembles of events for time averages. In addition, the short time history complicates obtaining the separable signal model (equation 4) that underpins matched field processing. Our solution is to break the signal into a large number of narrow frequency bands efficiently [see e.g. Portnoff, 1980], choose processing parameters that make signals in the narrow bands approximately independent, pursue matched field processing band by band and combine results incoherently across bands.

Narrowband signal representation

Figure 2 illustrates the type of seismic transient signals we consider, consisting of short forcing functions convolved with much longer Greens functions. To adapt matched field processing to transients of this sort, we break the received signal into a large number of narrowband components □□□, □ an integer band index □∈□□□□,⋯,□□□□. Each component is represented by the filtering operation:

□□□= □□ h□□−□ □□ (9)

We require that the bank of filters, represented by the collection of impulse responses h□□, allow perfect reconstruction of the signal from its narrowband components (a fidelity constraint):

□□= □□□□ (10)

One filterbank that satisfies this requirement is obtained by frequency translations of an ideal lowpass filter with impulse response □□ [Portnoff, 1980]:

ℎ□□= □□Δ□□□□□□

□□= 1Δ□ □□□□□□Δ□;   Δ□=2□Δ□   (11)

Φ□ =   1   □ ≤ Δ□2 0   □ > Δ□2

Figure 3 depicts impulse responses and corresponding frequency responses □□□ for the realizable digital filters we use to approximate this choice. The filterbank divides the frequency axis into disjoint bands of width Δ□ (Δ□=2□Δ□). We choose the bandwidth Δ□ to be small enough that the Fourier transform of the transient seismic source is approximately constant within any individual band. In the time domain this assumption corresponds to a requirement that the impulse response of the filters be large compared to the duration of the source. Our choice of bandwidth is Δ□= 40/128 = 0.3125 Hz. The duration of the impulse response, then, is 3 to 4 seconds, which is *usually* long compared to the durations of the explosions in Figure 2. In the figure, the signals are observed at a distance of about one kilometer. The source durations consequently are shorter than these observations.

The narrowband signal components □□□ are complex waveforms that consist of relatively slowly-varying complex envelope functions □□□ modulated by a complex exponential function [Franks, 1981]. This fact is readily apparent by substituting the impulse response of equation 11 into equation 9:

□□□ = □□Δ□□□□□□ □□−□ □−□Δ□□□ □□ = □□Δ□□□ □□□,   (12)

which resembles equation 3.

The signal component of the observations, the source time history and Greens functions also have representations in terms of narrowband complex envelopes, □□□□, □□□ and □□□□ respectively. Since the bands are disjoint, the convolution of equation 2 can be shown to obtain in each band individually:

□□□□= □□ □□□ □□□□−□ (13)

To support matched field processing, we seek a signal representation analogous to equation 4. This objective is aided by the short duration of the source. Since the complex envelopes are band-limited, we can invoke the sampling theorem [Oppenheim and Schafer, 1975] to represent the slowly-varying complex envelopes with discrete samples taken every Δ□ seconds. For example:

□□□= □ = −∞∞□□□ Δ□ □□□□□−□ Δ□

= Δ□□ = −∞∞□□□ □□−□ Δ□ (14)

Equation 14 represents the complex envelope as a series expansion in terms of sinc functions, where the coefficients in the expansion happen to be samples of the envelope □□□ =□□□ Δ□. This particular expansion has the desirable property that convolution among waveforms maps to convolution among the representative samples [Oppenheim and Schafer, 1975]:

□□□□ =□ = −∞∞ □□□ □□□□−□ (15)

Now, since the source time function has a short duration, it is approximately true that the envelope of the source is dominated by a single sample, chosen for convenience at □$= 0$:

□□□ ≈ 0; □ $\neq 0$ (16)

Substituting 16 into 15:

□□□□ ≈ □□*0* □□□□ (17)

consequently:

□□□□ ≈ □□*0* □□□□ (18)

which is the result we seek. We have expressed the signal in each narrow band as a separable product of the source time history and the Greens function. This approximation allows us to propose a probability model for the observations in which the second moment is structured in a manner similar to the outer product of equation 8. The principal difference between the seismic and underwater sound situations is that now the forcing functions appear as constants and the Greens functions are time-varying.

Probability model for the data

Assuming the source time histories are zero mean and normally distributed, the signals □□□ are zero-mean, non-stationary, Gaussian random processes, as are the complex envelopes □□□□ [Van Trees, 1968]. The probability distributions of the complex envelopes are characterized by their second moments:

□□□□□*1*, □*2* = □ □□□□*1* □□□□□*2* ≈ □□□*0*□□*∗0* □□□□*1* □□□□□*2* (19)

In contrast to the underwater sound case, these covariance functions depend on observation times, consistent with the non-stationary nature of the seismic signals. However, we concern ourselves with just the first P arrival (Pn) in this paper, so the covariance matrix of interest is limited to the quantity of equation 19 sampled at the P arrival time, i.e. at □*1*=□*2*=□□. We simplify the description of covariance by defining □□□= □□□0□□*0 and suppressing the dependence on □□:

□□□□ = □□□□□□, □□ ≈ □□□ □□□ □□□□ (20)

The matched field processing algorithm is considerably simplified if the narrowband signal components can be considered statistically independent, i.e. if □□□≙□□□ □□2, then:

□□□□ ≈ □□□ □□□; □□□= □□2 □□□ □□□□ (21)

where □□□ is the Kronecker delta function. We have found this condition to obtain approximately for mining explosions where there is significant variation in the source time history, as we will demonstrate in the next section. We summarize the observed data with the vector

□= □□□□□ □□□□□+1⁝ □□□□□ (23)

which is understood to consist of the array complex envelope functions in the bands ranging from □□□□ to □□□□ sampled at time □=□□. The probability density function for these data we take to be the complex multivariate normal density

□□= □ =□□□□□□□□□□−1□−□□□□□−1 □□ (24)

where

□□= □□2 □□□ □□□□ + □□ (25)

□□ is the covariance contributed by the ambient noise, still considered to be stationary. The assumption of independence among bands leads to a joint pdf which is the product of individual densities for the observations in each band.

Matched field processing algorithm

We use a matched field processing algorithm developed under simplifying assumptions of a high signal-to-noise ratio and noise which is statistically independent among frequency bands and channels of the array. The problem is to determine which of several possible (equally-likely) sources □ $\in \Theta$ is responsible for the observed signals □. If there are □ possible sources, this problem can be treated as an *M*-ary hypothesis test and is solved by maximizing the log likelihood function [Van Trees, 1968]:

□□□□= −□ = □□□□□□□□□□□□□□+□□□□□−*1* □□ (26)

over source □ . The solution to this problem is to choose the □ which maximizes the energy in the Pn window of a beam similar to the average power expressed in equation 7:

□∗= max□□ = □□□□□□□□□□□□□ □□□□□ □□□

□□□= □□□□□□□ □□□*12* (27)

We normalize the statistic in (27) to range between 0 and 1, by dividing the beam energy by the total energy incident on the array aperture:

□ = □□□□□□□□□□□□□ □□□□□ □□□□ = □□□□□□□□□□□□ □□ (28)

The modified statistic represents the energy in the Pn window representable or "captured" by the matched field steering vectors. We refer to (28) as the energy capture of the algorithm.

Calibration

The matched field processing algorithm of equation (27) requires knowledge of the vector of Greens functions □□□ in each frequency band for each source □ at least within a multiplicative constant. As in the underwater sound case, an estimate of □□□ can be obtained empirically from an estimate of the signal covariance □□□. This we obtain in practice from an estimate of □□ (equation 25) made from an ensemble of events from a particular source as described in the next section. The events are chosen to have sufficient high signal-to-noise ratio that the noise contribution □□ in equation 25 is negligible. Then an estimate of □□□ is obtained as the principal eigenvector of □□**.**

Data and Analytical Methods

As described in the introduction, we test the concepts of matched field processing and contrast it with conventional FK analysis using ARCES array observations of explosions in the Kola Peninsula region of northwest Russia (Figure 1). For our analysis, we selected 549 events attributed to the specific mines shown in Table 1 by reports from the mine operators, validated with data from stations (APA, LVZ) local to the mines. For each event, we acquired 200 seconds of waveform data (Figure 2) for the □□*=17* ARCES elements comprising the center and the outer two rings of the ARCES array (aperture approximately 3 kilometers), and manually picked the Pn onsets. We note that at this range, the Pn phase is well separated from Pg and the later phases, so that we can consider processing Pn as a temporally isolated branch of propagation, even in the narrow frequency bands that we choose for our analysis.

Estimation of covariance matrices

The principal data reduction we performed was to compute estimates of covariance matrices □□ in each band □ □ □□□□,⋯, □□□□. There were two types of covariance matrix estimates: those for single events and ensemble averages (Table 1, Figure 4) for each mine. To estimate these covariance matrices, we assembled the event observations □□□; □=1,⋯,□□ from a particular source □, aligned the waveforms to the first P arrival, filtered them into their constituent narrow bands □□□□□, then computed

□□□ = 1□Δ □=0□Δ−1□□□□0+□Δ□□□□0+□Δ□ (29)

for the individual covariance matrices, where □0 was the pick time, Δ=0.125 seconds, □Δ=32, and

□□= 1□□ □=1□□□□□□□□□ (30)

for the ensemble averages. We performed averaging over 4-second windows of the complex envelopes in equation 29, to ensure that the sampling of Pn correlation characteristics was representative of the bulk of Pn energy. The arrival time of the energy varies from event to event. For the purposes of this investigation □□□ ≈ □□□□□□□□□□ since the window selected contained only the Pn phase in all cases (Figure 4). The individual event contributions to the ensemble average in equation 30 were normalized by the total energy to prevent a few larger events from dominating the average. We also required a signal-to-noise ratio greater than 2 in each band, so that the ensemble covariance would be dominated by signal characteristics, i.e. □□ ≈ □□□.

The frequency bands also were selected for good signal-to-noise ratio over the collection of 549 events. The processing band ranged from 2.5 Hz (□□□□ = 8) to 12.5 Hz (□□□□ = 40), with the width of the bands chosen to be *Δ*□ = 0.3125 Hz. Consequently, the data were processed in 33 bands as shown in Figure 3.

Assumption of independence among frequency bands

We performed a separate calculation of the covariance among signals in different frequency bands as a check of assumed independence. Equations 29 and 30 were modified to correlate □□□□ and □□□□, envelopes in different frequency bands, to estimate □□□□ for □,□ ∈[8,⋯,40]. We then constructed the matrix

□ = □8 8⋯□8 40⋮⋱⋮□40 8⋯□40 40 = □ □ □□ (31)

and normalized the rows and columns of the matrix by the square root of the product of the diagonal elements they contain. This normalization leaves a diagonal consisting of all ones; the off-diagonal elements represent the coherences between the signals observed by individual array sensors in frequency bands □ and □. The resulting matrix is shown in Figure 5 and is approximately block diagonal, where each block is a 17×17 submatrix representing the coherence among the □□=17 ARCES elements used in the calculation. Since the off-diagonal elements are complex, the magnitude of the elements is rendered as an image, with bright pixels representing high coherence (approaching 1) and dark pixels representing low coherence (approaching 0). The low coherence indicated by the off-diagonal blocks suggests that the frequency bands are approximately uncorrelated, supporting the approximation of equation 21.

Characteristics of steering vectors

The successful and routine use of FK analysis motivates a comparison of matched field processing to FK methods. The frequency-wavenumber spectrum has the same mathematical structure as the matched field processing statistic, but with steering vectors defined by a plane wave model (see e.g. Aki and Richards, 1980, section 11.4). The FK spectrum [illegible] maps power (or energy) incident upon an array as a function of frequency [illegible] and wavenumber [illegible], and this spectrum is estimated by:

[illegible] = [illegible] (32)

where [illegible] is the estimated spectral covariance matrix of the observations evaluated at frequency [illegible], and the steering vectors have the form:

[illegible] = [illegible] (33)

The FK equivalent of equation 27 is:

[illegible] = max [illegible] = [illegible] (34)

where [illegible] is the slowness vector defined by the nominal Pn velocity in the region and the great circle back-azimuth from the array to the source [illegible]. We note that the statistic in (34) is a wideband extension of the single-frequency FK definition that assumes uncorrelated frequency components and is frequently used in practical array operations [Kvaerna and Doornbos, 1986; Kennett, 2002].

It is common to calibrate FK analysis by introducing corrections for vector slowness estimated from (34) over the observation band using collections of ground truth events (see e.g. Schweitzer, 2001; Gibbons et al., 2009). Gibbons et al. (2009) in particular demonstrate that calibrations of refraction can improve the ability of small-aperture arrays to distinguish events from closely-spaced mines. A fair comparison of FK analysis and matched field processing should include FK slowness corrections. We estimate slowness corrections in our analysis by selecting the slowness to maximize the FK energy estimate for a particular source using the ensemble covariance matrices estimated for that source:

□□∗= max□□ = □□□□□□□□□□□□□□□ □□□ (35)

Subsequently we use s□∗ instead of the slowness vector predicted by the great circle path to generate plane wave steering vectors, when we speak of calibrated FK analysis.

It is instructive to visualize the plane wave steering vector and its empirical matched field counterparts to understand why matched field processing performs so well. It is reasonable to expect that, since FK analysis generally is successful at low frequencies, the empirically-derived steering vectors should reproduce their plane-wave counterparts in the frequency band where arrays are considered coherent. This expectation is largely fulfilled, as we will show. However, the empirical steering vectors depart dramatically from their plane wave counterparts as frequency increases.

Figure 6 introduces our visualization approach. The □□ elements of the steering vectors are complex phasors characterized by a magnitude and a phase. In the figure, each phasor is rendered as a circular symbol at the map position of the corresponding physical array element, with symbol size proportional to the phasor magnitude and color keyed to the phase. In the plane wave model, the phasors have uniform magnitude and the phase is obtained by projecting the vector representing the array element location onto the slowness vector (Figure 6 left). In the plane wave case phase is constant in planes perpendicular to the direction of wave travel and the phasor magnitude is constant (Figure 6 right).

Figure 7 presents a panel of plane-wave steering vectors rendered as described in Figure 6. This panel compares the steering vectors for each of the ten mines in each of seven different frequency bands (⍰ $\in [10, 15, 20, 25, 30, 35, 40]$), which are the bands highlighted in blue in Figure 3. The steering vectors have been optimized by the slowness calibration of equation 35. The principal lesson of Figure 7 is that, even with slowness corrections, the steering vectors do not differ very much among the mines. Variation among the steering vectors is negligible for the lowest frequency (3.125 Hz) displayed, but does not increase very much as frequency increases even to 12.5 Hz. At 12.5 Hz there is noticeable variation between the Olenegorsk and Khibiny groups but minor variations among individual mines within those two groups. This observation suggests that FK analysis of the Pn phase, even with calibrations, will not provide a reliable means to distinguish events from these ten sources.

The empirical steering vectors are estimated as the principal eigenvectors of ensemble covariance matrices (equations 29 and 30) for each of the mines. This calibration approach assumes that the eigenspectra of the covariance matrices are dominated by a single eigenvalue, so that the covariance matrix is well approximated by the outer product of the associated principal eigenvector with itself (equation 19). Figure 8 shows the proportion of energy in the eigenspectrum concentrated in the largest eigenvalue for each of the ten mines as a function of frequency band. Below 7 or 8 Hz a single eigenvalue does dominate. For the Olenegorsk group of mines, the largest eigenvalue accounts for at least 60% of the energy even to 12.5 Hz. The

Khibiny group shows a different behavior, with a sharp drop in concentration of energy in the largest eigenvalue above 7 Hz. But the largest eigenvalue still accounts for at least 40% of the energy at 12.5 Hz. The drop in concentration of the eigenspectrum energy is evidence of some heterogeneity in the sources (mines) of the Khibiny group (i.e. they do not behave as simple point sources). The Kirovsk and Rasvumchorr explosions, for example, are designed to drop the roofs of mining drifts with fans of charges emplaced in a series of holes radiating from the drifts like spokes in a wheel. Several such fans may be detonated within a few tenths of a second.

Steering vectors comparable to those displayed in Figure 7, but obtained empirically from the estimated ensemble covariance matrices are displayed in Figure 9. Two features of these data are notable: there is a great deal more variability among the empirical steering vectors than among their theoretical plane wave counterparts and the amplitudes of elements comprising the vectors are significantly non-uniform. At 3.125 Hz, the steering vectors are largely homogeneous among all of the mines, indicating a desirable reproducibility of our data reduction techniques and the consistency one would anticipate at low frequencies where the effects of scattering are less pronounced. As the frequency increases, the variation among the mines increases dramatically. In the next two higher bands displayed (4.6875 and 6.25 Hz) the two mine groups are individually homogenous, but variations between groups are significant. As the frequency increases above 7 Hz, variations among the mines within each group become increasingly significant until the phase patterns appear almost random at 12.5 Hz. The high degree of heterogeneity in these steering vectors suggests that empirical matched field processing should perform well to resolve events originating at these mines. The degree of variation in magnitude from sensor to sensor is notable, suggesting focusing effects in wave propagation.

One way to determine whether empirical steering vectors reproduce plane wave steering vectors is to examine the inner product between empirical and plane wave (theoretical) vectors:

$$[illegible] \quad (36)$$

Since both empirical and theoretical steering vectors have been normalized to unit length, the magnitude of the inner product must range between 0 and 1, approaching 1 if the vectors are similar. Figure 10 shows this comparison for the ten mines, with generally high similarity between theoretical and empirical steering vectors below 7 Hz, and low similarity above that frequency. The two groups of mines behave quite differently. The variability in the inner product is remarkably small among the Khibiny mines below 7 Hz but large among the Olenegorsk group. In general, however, the calibration process for steering vectors does tend to reproduce plane wave vectors at low frequencies where regional arrays are considered to operate coherently and where FK analysis typically is employed successfully.

The disparities among steering vectors shown in Figures 7 and 9 can be made more apparent by examining the relative amplitude and phase for each array element from mine to mine. In Figures 11 and 12, we select one mine (K2: Rasvumchorr) as a reference and display the elements [illegible], [illegible]$=1,\cdots,$ [illegible] of steering vectors normalized by the corresponding element for the K2 vector, i.e.

$$[illegible] \quad (37)$$

Figure 11 shows the normalized plane wave steering vectors and Figure 12 the normalized empirical steering vectors. Note that in both figures the normalized steering vectors for

Rasvumchorr have elements with uniform magnitude and zero phase, as the reference should. In Figure 11, steering vectors for the remainder of the mines show small phase variations at low frequencies that increase steadily as frequency increases. The Khibiny mines are all very similar to Rasvumchorr. In Figure 12, the variations are much more dramatic and include big variations in amplitude as well as phase. As in Figure 11, the variations increase with increasing frequency. However, at the highest frequencies, the Khibiny mines show as much variation as the Olenegorsk mines.

We interpret these results as the effects of scattering, which become more pronounced as frequency increases. We speculate that the scale length of heterogeneities affecting scattering becomes significantly less than the separations between mines in the two groups at the higher frequencies analyzed. Ultimately variations between mines within one of the two groups can become as large as variations between (the more distant) groups.

It is instructive to examine another measure of geographic resolution offered by the processing techniques. The ambiguity function [Baggeroer et al. 1993]

□□*1*, □*2*= □□*1*□□□*2*□ (38)

measures the degree of similarity between steering vectors. For perfectly resolved sources □□*1*, □*2*= □□*1*□*2*. To obtain a sense of geographic resolution, a reference point can be selected, for example □*1*= □∗ and the ambiguity function can be mapped against other geographic points □*2*. The theoretical ambiguity function assuming plane wave propagation is contoured at three different frequencies in Figure 1 using Rasvumchorr as the reference. Figure 13 plots the ambiguity functions evaluated at each of the ten mines again using Rasvumchorr as the reference. The value of the ambiguity function is represented as a symbol at the geographic locations of the mines: symbol size is proportional to the ambiguity value. A matrix of six cases is shown: 2 different frequencies (5 and 10 Hz) and three different types of steering vectors. The first type (left) is the theoretical plane wave steering vector assuming slowness vectors determined from the great circle path from mine to the array and a nominal Pn velocity (8.0 km/sec) for the region. The second type (middle) is the plane wave steering vector with slowness corrections calibrated by the method of equation 35. The third type (right) is the matched field steering vector. At 5 Hz, only the matched field steering vectors appear to reject the Olenegorsk mines very well and none of the methods provides very unambiguous separation of Rasvumchorr from the other Khibiny mines. At 10 Hz, only the matched field method clearly rejects the other Khibiny mines.

Event classification results

Our principal objective is to use matched field processing (equation 27) to determine the origins of events from the Olenegorsk and Khibiny mining regions using observations of the Pn phase with the ARCES array. We contrast the performance of this method with FK analysis (equation 35) with and without slowness corrections. For the two methods that require calibrations, we used a cross validation approach to avoid problems of circularity. We held one of the 549 events out, calibrated steering vectors using the remaining 548 events then used the calibrated steering vectors to classify the one reserved event. We iterated this procedure 549 times using each event in turn as the reserved event.

It is instructive to examine histograms of the classification statistics, i.e. the matched field or FK energy computed in the 33 bands ranging from 2.5 to 12.5 Hz. These statistics are normalized in the manner of equation 28 to represent energy capture for each of the processing methods. Figure 14 summarizes these histograms for the 52 explosions of the Olenegorsk mine (O2). The figure shows 10 histograms for each of the three processing options. Each of the 10 histograms corresponds to a particular hypothesis about the mine of origin. The leftmost column of the figure shows the energy capture distributions for FK analysis using plane wave steering vectors constructed from theoretical (great circle) back azimuths. There is very little difference among the distributions, which are clustered around 0.2-0.3 energy capture, except that the distributions for the 5 Olenegorsk mine hypotheses are shifted to slightly higher values than those of the corresponding 5 Khibiny hypotheses.

The center column of the figure presents the results for the FK algorithm with slowness corrections. The corrections clearly improve the separation of the distributions under the Olenegorsk hypotheses from the Khibiny distributions. However, the distribution under the O2 hypothesis is not significantly distinct from the O1 and O3-O5 distributions.

The matched field processing results are summarized in the third column. With matched field calibrations the separation of the O2 distribution from the other distributions is dramatic.

The results of classifying the 549 events from all mines are summarized in Figure 15. This figure has three parts, one for each processing approach. The top part summarizes the classification results for the FK method with no corrections. It consists of 10 histograms, one stacked behind the other showing the frequency of classification for the events of each mine under each of the 10 hypotheses about event origin. A perfect result would show a diagonal of filled bins containing all of the events correctly attributed to their origins. For this method (labeled "theoretical plane wave"), the majority of events are incorrectly assigned to one mine, the Rasvumchorr mine (K2). We attribute this behavior to the fact that observed Pn backazimuths are biased clockwise (to the south) and the K2 mine is the southernmost of all the mines. Consequently, the steering vectors for this mine present the best fit to the refracted wavefields.

The middle part (labeled "empirical plane wave") summarizes results for the FK method with slowness corrections. The slowness corrections presumably remove gross biases in azimuth and the classification performance is significantly better. Few errors are made between the two mining groups, which is consistent with other results obtained with wideband FK analysis [Gibbons et al., 2009]. Within the Olenegorsk group, classification results are correct more often than not. But classification performance for mines within the more distant Khibiny group is not good, representing a more nearly uniform (random) assignment of events to originating mines. We interpret this result to indicate that the separations of these mines are below the resolution limit of the array, even at 12.5 Hz, and that coherent processing based on a plane wave model is not possible in the higher frequency bands.

The bottom part (labeled “matched field”) summarizes the matched field processing classification results. Matched field processing identifies the sources of the explosions with a high degree of reliability (98.2%, 539 of 549 events correctly identified).

Conclusions

Matched field processing with empirically-calibrated steering vectors reliably identifies the origins of explosions in the Khibiny and Olenegorsk mines using just observations of the Pn phase made with the ARCES array. This result is remarkable because conventional considerations of array resolution (Rayleigh limit) suggest that poor performance should be expected. Indeed our comparison with a conventional broadband FK algorithm, even with slowness corrections, indicates that algorithms that rely upon a plane wave model perform in a manner consistent with expectations. The matched field results appear to be a consequence of the rich scattering environment for wave propagation in the crust and upper mantle. The mines are too closely spaced (just tenths of a degree apart in azimuth) to be resolved under free space propagation conditions, which would offer only direct-path propagation. The highly heterogeneous environment creates multipath propagation, which produces an apparently disorganized phase and amplitude structure across the array aperture (Figures 9 and 12) that may be calibrated with observations of previously-recorded events. This structure appears to be repeatable from event to event, allowing an algorithm which exploits that structure to identify closely-spaced sources reliably.

We speculate that the explosions in the Kola peninsula may illuminate a field of scatters over a much larger angle, when observed by ARCES, than that subtended by the mining region itself (~ 5 degrees), and perhaps larger than the classical resolution limit. We suggest that explosions at the 10 mines illuminate the field of scatterers in unique patterns, depending on their positions within that field, and that the seismic wavefields scattered toward the array across the broad scattering aperture differ significantly among the mines. This interpretation is consistent with techniques in cellular telephony widely used to increase communication bandwidth with antenna arrays in a rich scattering environment [Foschini and Gans, 1998; Simon et al., 2001; Moustakas et al., 2000]. Empirical matched field calibration captures the detail of refracted and scattered wavefield structure across the array aperture, leading to superior performance in distinguishing the sources. Since mining explosions are distributed (ripple-fired) sources, source radiation pattern effects also may make a contribution to distinguishing the fields as energy is directed outward in a unique azimuthal pattern from each mine.

Wideband matched field processing processes array observations incoherently across a large number of narrow frequency bands, which confers a degree of immunity to variations in source time history. Waveform correlation methods, a potential competing solution for the kind of classification problem we have discussed, are, by contrast, more sensitive to variations in the source signal. We make this statement based upon experience with both algorithms, though a formal study needs to be done to quantify these effects.

However, like waveform correlation methods [e.g. Gibbons and Ringdal, 2006], matched field processing calibrates the spatial structure of a signal across the observing aperture, permitting coherent processing at frequencies for which the signal is considered incoherent under a plane wave model. This observation suggests that there may be no limit to the size of the aperture over which coherent processing may be attempted. Observing networks that were deployed with only incoherent processing methods in mind might now be used in a coherent processing framework.

The basic technique we describe in this paper can be extended in several useful ways. The assumption that the source must be idealized as a single point force can be relaxed to allow more heterogeneous sources modeled by a number of independent point sources. We had a suggestion in our data (Figure 8) that this extension is desirable. The concentration of eigenspectrum energy in the top eigenvalue declines at high frequencies for the mines of the Khibiny group. The SNR of events is high for this group above 7-8 Hz, suggesting that some factor other than poor SNR is at play. This factor could be source heterogeneity. Changing the signal model to acknowledge source heterogeneity requires the signal covariance matrices [illegible] to have ranks greater than one. The matched field processing approach can be modified to use a collection of steering vectors extracted as the now multiple principal eigenvectors of the ensemble covariance matrix in the calibration step. The algorithm must be modified to maximize power over a subspace of steering vectors spanned by the collection of principal eigenvectors. This observation provides additional justification for the use of event ensembles for estimating covariance matrices, as the rank of an estimated matrix cannot exceed the number of events used in the estimate. The number of effective point sources comprising a heterogeneous source can be discovered through eigendecomposition of an ensemble covariance matrix.

The other principal extension that would be useful is to modify matched field processing to use more of the waveform than just the first arriving P phase, extending perhaps to the entire seismogram. This modification would entail use of the entire covariance function and its eigendecomposition in a Kahunen-Loeve (KL) expansion [Van Trees, 1968] to develop matching fields that capture and use the time evolution of spatial structure. The narrowband decomposition still would be required to suppress the effects of a short, variable source time history, and a KL expansion would be required in each frequency band. While complicated, this approach is feasible with modern computing capabilities.

References

Aki, K. and P. G. Richards, 1980, *Quantitative Seismology*, Vol. 2, W. H. Freeman and Co., San Francisco.

Baggeroer, A. B., Kuperman, W. A., and Mikhalevsky, P. N., 1993, An overview of matched field methods in ocean acoustics*, IEEE Journal of Oceanic Eng*ineering, vol. 18, number 4, 401-424.

Bucker, H. P., 1976, Use of calculated sound fields and matched field detection to locate sound sources in shallow water, *Journal of the Acoustical Society of America,* volume 59, number 2, 368-373.

Capon, J., 1969, High resolution frequency-wavenumber spectrum analysis, *Proceedings of the IEEE,* volume 57, number 8, 1408-1418.

Fialkowski, L. T., Collins, M. D., Kuperman, W. A., Perkins, J. S., Kelly, L. J., Larsson, A., Fawcett, J. A., and Hall, L. H., 2000, Matched-field processing using measured replica fields, *Journal of the Acoustical Society of America,* volume 107, number 2, 739-746.

Foschini, G. J. and M. J. Gans, 1998, On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications* **6** (3): 311–335.

Franks, L. E., 1981, *Signal Theory*, Dowden and Culver, Inc., Stroudsburg, PA, 79-96.

Gibbons, S. J. and Ringdal, F., 2006, The detection of low magnitude seismic events using array-based waveform correlation, *Geophysical Journal International,* volume 165, 149-166.

Gibbons, S. J., Sorensen, M., Harris, D. B., Ringdal, F., 2007, The detection and location of low magnitude earthquakes in northern Norway using multichannel waveform correlation, *Physics of the Earth and Planetary Interiors*, volume 150, number 3-4, 285-309.

Gibbons, S. J., T. Kvaerna and F. Ringdal, 2009, Considerations in phase estimation and event location using small aperture regional seismic arrays, *Pure and Applied Geophysics*, accepted for publication

Harris, D. B., 1991, A waveform correlation method for identifying quarry explosions, *Bulletin of the Seismological Society of America,* volume 81, 2395–2418.

Hauksson, E. and Shearer, P., 2005, Southern California hypocenter relocation with waveform cross-correlation, Part 1: Results using the double-difference method*, Bulletin of the Seismological Society of America,* volume 95, 896-903.

Israelsson, H., 1990, Correlation of waveforms from closely spaced regional events, *Bulletin of the Seismological Society of America,* volume 80, 2177-2193.

Johnson, D. H. and Dudgeon, D. E., 1993, *Array Signal Processing: Concepts and Techniques*, Prentice Hall.

Kennett, B.L.N., 2002, *The Seismic Wavefield. Volume II: Interpretation of Seismograms on Regional and Global Scales*, Cambridge University Press.

Kværna, T. and Doornbos,D.J, 1986, An integrated approach to slowness analysis with arrays and three-component stations, NORSAR Scientific Report: Semiannual Technical Summary No. 1-1986/1987, NORSAR, Kjeller, Norway. pp. 29-40.

Moustakas A. L., H. U. Baranger, A. M. Sengupta, L. Balents and S. H. Simon, 2000, Communication through diffusive media: coherence and capacity, Science, vol. 287, p. 287 .

Mykkeltveit, S., Aastebol, K., Doornbos, D. J., and Husebye, E. S., 1983, Seismic Array Configuration Optimization, *Bulletin of the Seismological Society of America.* 73**,** 173-186.

Oppenheim, A. V. and R. W. Schafer, 1975, *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey.

Portnoff, M. R., 1980,Time-frequency representation of digital signals and systems based on short-time Fourier analysis, *IEEE Trans. on Acoustics, Speech and Signal Processing*, volume ASSP-28, number 1, 55-69.

Schaff, D. P., and Richards, P. G., 2004, Repeating seismic events in China, *Science*, volume 303, 1176-1178.

Schaff, D. P. and Waldhauser, F., 2005, Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bulletin of the Seismological Society of America,* volume 95, 2446–2461.

Schweitzer, J., 2001, Slowness corrections – one way to improve IDC products, *Pure and Applied Geophysics,* volume 158, 375-396.

Simon S. H., A. L. Moustakas, M. Stoytchev and H. Safar, 2001, Communication in a disordered world, *Physics Today*, pp 38-43.

Van Trees, H. L., 1968, *Detection, Estimation, and Modulation Theory, Part I.* Wiley, NewYork.

Van Trees, H. L., 2002, *Detection, Estimation, and Modulation Theory, Part IV*: *Optimum Array Processing*, Wiley-Interscience, New York.

Waldhauser, F. and Ellsworth, W. L., 2000, A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California, *Bulletin of the Seismological Society of America,* volume 90, 1353-1368.

Acknowledgements

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 and NORSAR under contract DE-FC52-05NA26604. The authors thank Steve Myers, Bill Walter and Frode Ringdal for their help and suggestions.

Figure Legends

Figure 1 The mines of the Khibiny and Olenegorsk regions are too close for explosions to be attributed to specific mines on the basis of wavenumber (FK) spectrum direction estimates made from ARCES array (upper left) observations of Pn waves (Figure 2). ARCES resolution is indicated on the map at lower left, which shows the half-power contours of the array response at three frequencies when the array is steered to the Rasvumchorr mine (K2).

Figure 2 Recordings of 7 explosions conducted at the Rasvumchorr mine (see Figure 1) made at an in-mine station (upper left) and the ARCES array center element (bottom, upper right) demonstrate that the regional seismic signal is the convolution of a short excitation and a prolonged Green's function. The great variability of the signals seen near the source and the common path to the observing station suggest the array signal can be modeled as the convolution of brief stochastic forcing functions with long-duration, deterministic Green's functions. Consequently, the spatial structure of the signal is deterministic (but unknown), even if the time history is a random process.

Figure 3 The real parts of the first 10 impulse responses (top) and amplitude frequency responses (bottom) for the bank of narrowband filters defined in equation (11). Seven individual bands are highlighted in blue and the sum of 33 bands in the range 2.5-12.5 Hz is shown in red.

Figure 4 Signals recorded by the ARCES station ARA0 (the array center element) for the ensemble of events used to calibrate the Norpakh mine (K5). Note the first-arriving P waves (see inset) show a high degree of time history variation.

Figure 5 This figure shows the matrix of inter-element and inter-band coherence of the Pn phase for the 17-element subarray of the ARCES shown in Figure 1. The coherence has been computed here for explosions at the Kirovsk mine, using 114 events. The image consists of 33×33 blocks each representing a distinct pairing of 33 different frequency bands (the labels on the axes represent frequency). Each block is a 17×17 matrix of correlations among the elements of the ARCES array. The fact that the matrix is approximately block diagonal is evidence that the 33 frequency bands chosen for our analysis are approximately uncorrelated.

Figure 6 This figure introduces a method for visualizing steering vectors that is used subsequently in Figures 7, 9, 11 and 12. The steering vector (right) can be visualized by plotting a symbol at each array sensor location. The size of the symbol is proportional to the magnitude of the steering vector element corresponding to the sensor. The color of the symbol encodes the phase. The example shown here is for a plane wave incident on the array from the southeast; the amplitudes are uniform and the phase increases linearly with distance along the direction of wave travel. The phase is calculated from the geometry shown at left: the phase is proportional to the vector offset of a sensor projected onto the slowness vector defining the direction and velocity of travel.

Figure 7 Plane wave steering vectors in 7 frequency bands (highlighted in blue in Figure 3) for all 10 mines are rendered here with the method described in Figure 6, and show that theoretical plane wave steering vectors vary little among the mines. The slowness for each mine has been optimized to fit the data (maximize the ensemble FK spectrum) over the 2.5-12.5 Hz frequency band (equation 36).

Figure 8 At all frequencies, the ensemble covariance matrices have a large fraction of their eigenspectrum energy concentrated in a single eigenvalue. The two plots show the fraction of eigenspectrum energy present in the largest eigenvalue as a function of frequency for the 5

Khibiny mines (top) and the 5 Olenegorsk mines (bottom). The light lines depict the concentration of eigenspectrum energy for individual mines and the dark lines depict mining group averages.

Figure 9 Empirical (matched field) steering vectors are depicted here in the same manner as in the Figure 7, allowing direct comparison. Note the variability among the mines (vertical direction) which increases at high frequencies. Steering vectors diverge substantially above 8 Hz.

Figure 10 Plots of the inner products (equation 37) of theoretical and empirical steering vectors show that, at low frequencies, the matched field calibration process largely reproduces the theoretical plane wave steering vectors. At high frequencies deviation increases demonstrating the breakdown of the plane wave model. The inner products for individual mines are shown as light lines; averages for each mining group are depicted with heavy lines. Mines of the Khibiny and Olenegorsk groups behave differently, but both show poor matches between measured and theoretical steering vectors above 8 Hz.

Figure 11 This figure shows how calibrated plane-wave steering vectors vary from a reference vector (choosing the Rasvumchorr mine (K2) as the reference). Symbol size represents amplitude difference and color phase difference between an element of a particular mine's steering vector and the corresponding element of the Rasvumchorr steering vector. The format for this plot is the same as that in Figures 7 and 9. Ten mines and 7 frequencies are depicted.

Figure 12 This figure portrays the differences among empirical (matched field) steering vectors comparable to the plane wave vectors of Figure 11. Note the very great differences here, increasing with frequency, suggesting much greater potential for differentiating signals among the mines.

Figure 13 This figure shows the increase in resolution obtained with empirical matched field steering vectors. The quantity displayed is the ambiguity function, the inner product between steering vectors at each mine and the corresponding reference steering vector at the Rasvumchorr mine. The ambiguity value is rendered as symbol size (area) at the geographic location of each mine. Ambiguity functions for three different types of steering vectors are shown at each of two frequencies: 5 and 10 Hz. The theoretical plane wave function results from steering vectors calculated from great circle path azimuths and a fixed phase velocity of 7.8 km/sec. The calibrated plane wave function results from fitting the best plane wave to the data over the 2.5-12.5 Hz band. The matched field function has much greater resolution which results from steering vectors extracted as the principal eigenvectors of covariance matrices for each mine in the two frequency bands.

Figure 14 Distributions of the matched field processing classification statistic (right) for the Olenegorsk (O2) mine population of 52 events under the 10 hypotheses about originating source (indicated at left) separate O2 unambiguously from the other mines. Frequency distributions for the theoretical (great-circle path) plane wave FK classifier (left) are ambiguous. Distributions for the plane wave FK classifier with slowness corrections (middle) show slightly improved separation between the two mining groups.

Figure 15 Classification results for 549 events show that the theoretical plane wave (FK without corrections) spectrum (top) is unable to resolve the mines, as expected. The FK algorithm with corrections (middle) largely separates the Olenegorsk group from the Khibiny group, and begins to distinguish individual mines. The matched field method (bottom) separates all mines with a high degree of reliability (98.2%).

Tables

Table 1 Number of events at each mine observed by the ARCES, together with distance and backazimuth to each mine

Mine (□)	ARCES Distance	ARCES Backazimuth	Number of explosions (□□)
Kirovogorsk	341	114.3	37
Olenegorsk	346	112.8	52
Oktjabrsk	347	114.3	18
Bauman	349	114.5	35
Komsomolsk	356	113.4	29
Kirovsk	393	118.0	114
Rasvumchorr	400	118.0	108
Central	403	118.0	63
Koashva	406	117.5	56
Norpakh	409	116.6	37

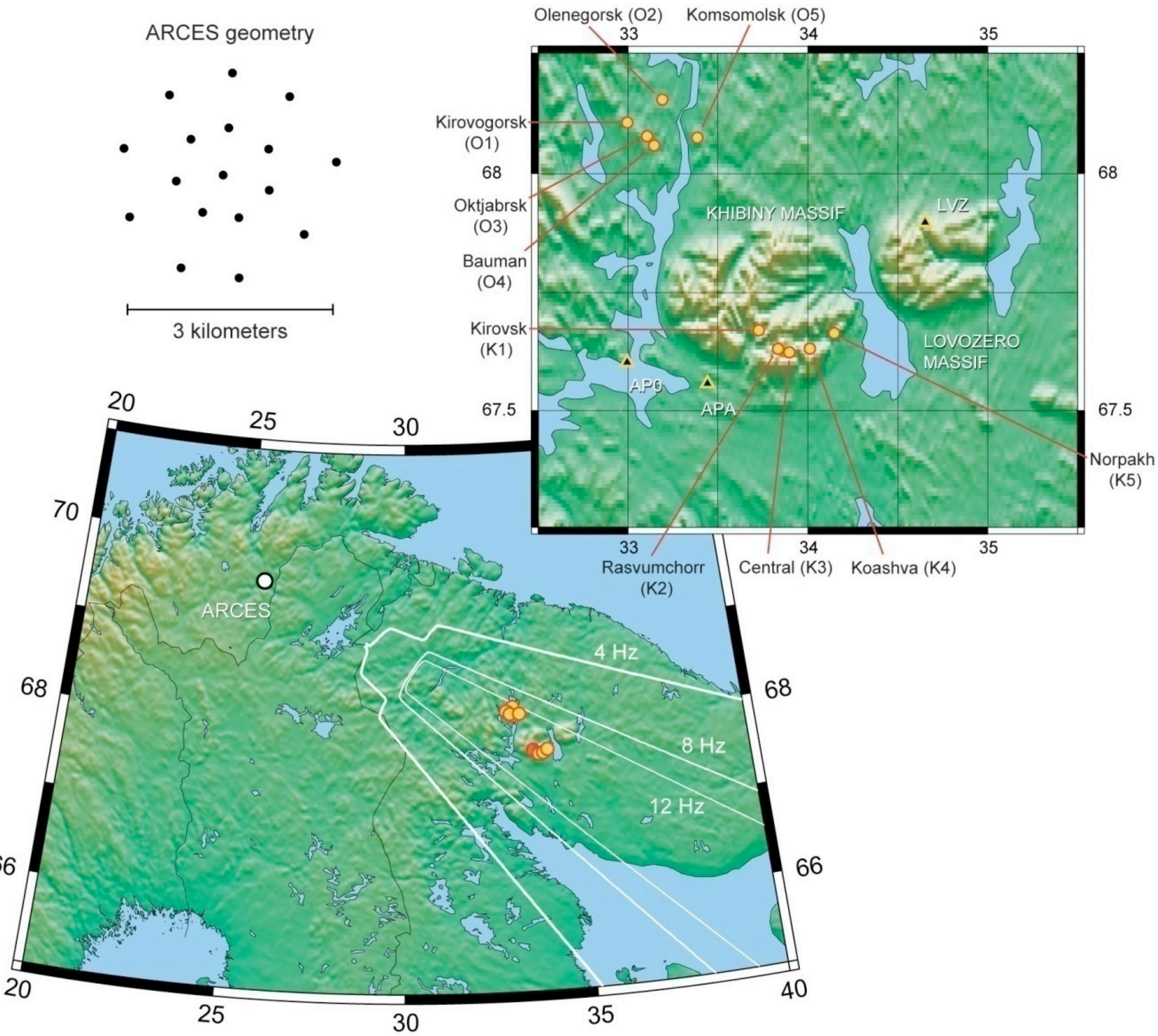

ARCES geometry
3 kilometers
Olenegorsk (O2)
Komsomolsk (O5)
Kirovogorsk (O1)
Oktjabrsk (O3)
Bauman (O4)
Kirovsk (K1)
KHIBINY MASSIF
LVZ
LOVOZERO MASSIF
AP0
APA
Norpakh (K5)
Rasvumchorr (K2)
Central (K3)
Koashva (K4)
ARCES
4 Hz
8 Hz
12 Hz

Figure 1

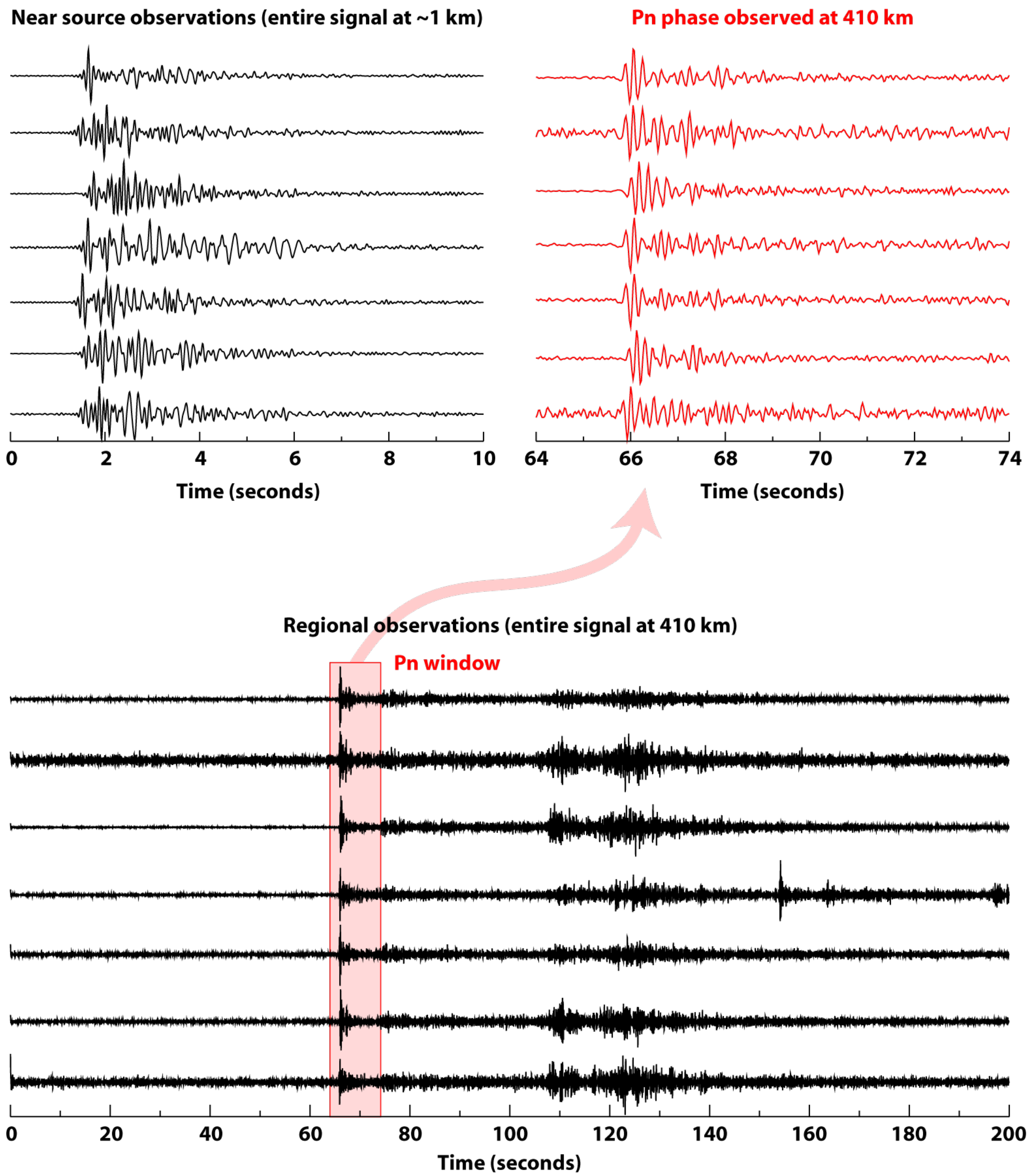
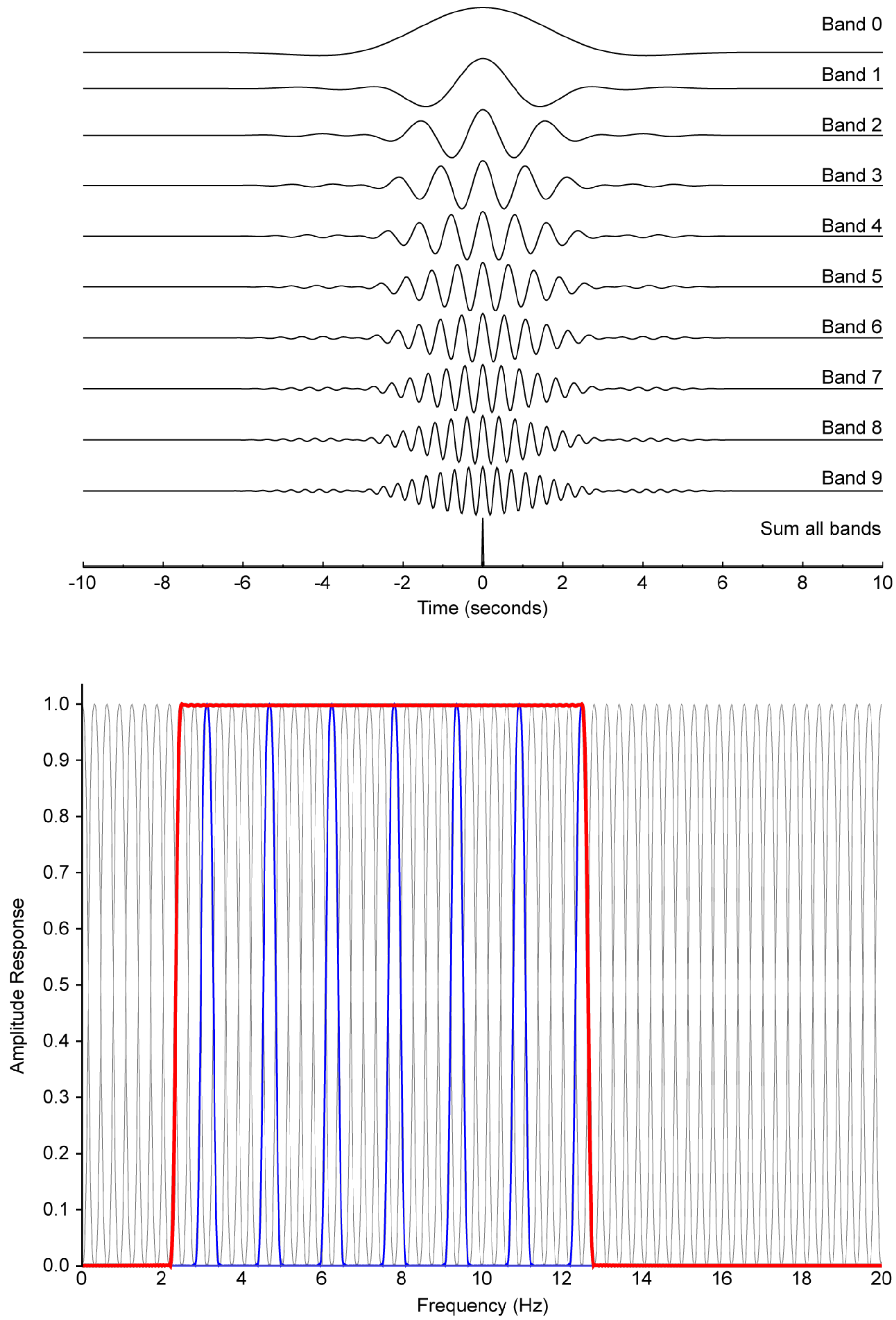
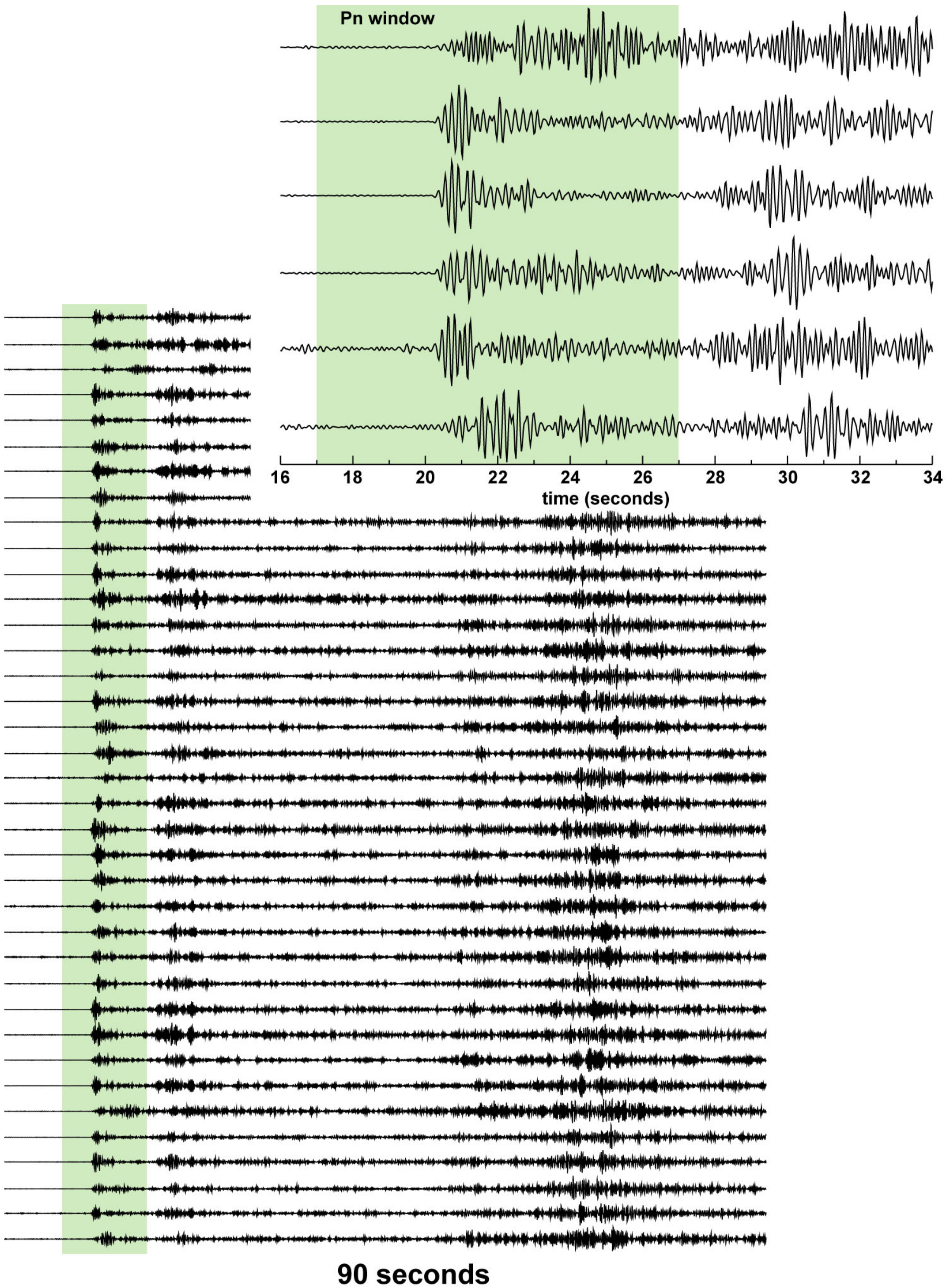

Near source observations (entire signal at ~1 km)
Pn phase observed at 410 km
0
2
4
6
8
10
Time (seconds)
64
66
68
70
72
74
Time (seconds)
Regional observations (entire signal at 410 km)
Pn window
0
20
40
60
80
100
120
140
160
180
200
Time (seconds)

Figure 2



Band 0
Band 1
Band 2
Band 3
Band 4
Band 5
Band 6
Band 7
Band 8
Band 9
Sum all bands
-10
-8
-6
-4
-2
0
2
4
6
8
10
Time (seconds)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Amplitude Response
0
2
4
6
8
10
12
14
16
18
20
Frequency (Hz)

Figure 3



Pn window
16
18
20
22
24
26
28
30
32
34
time (seconds)
90 seconds

Figure 4

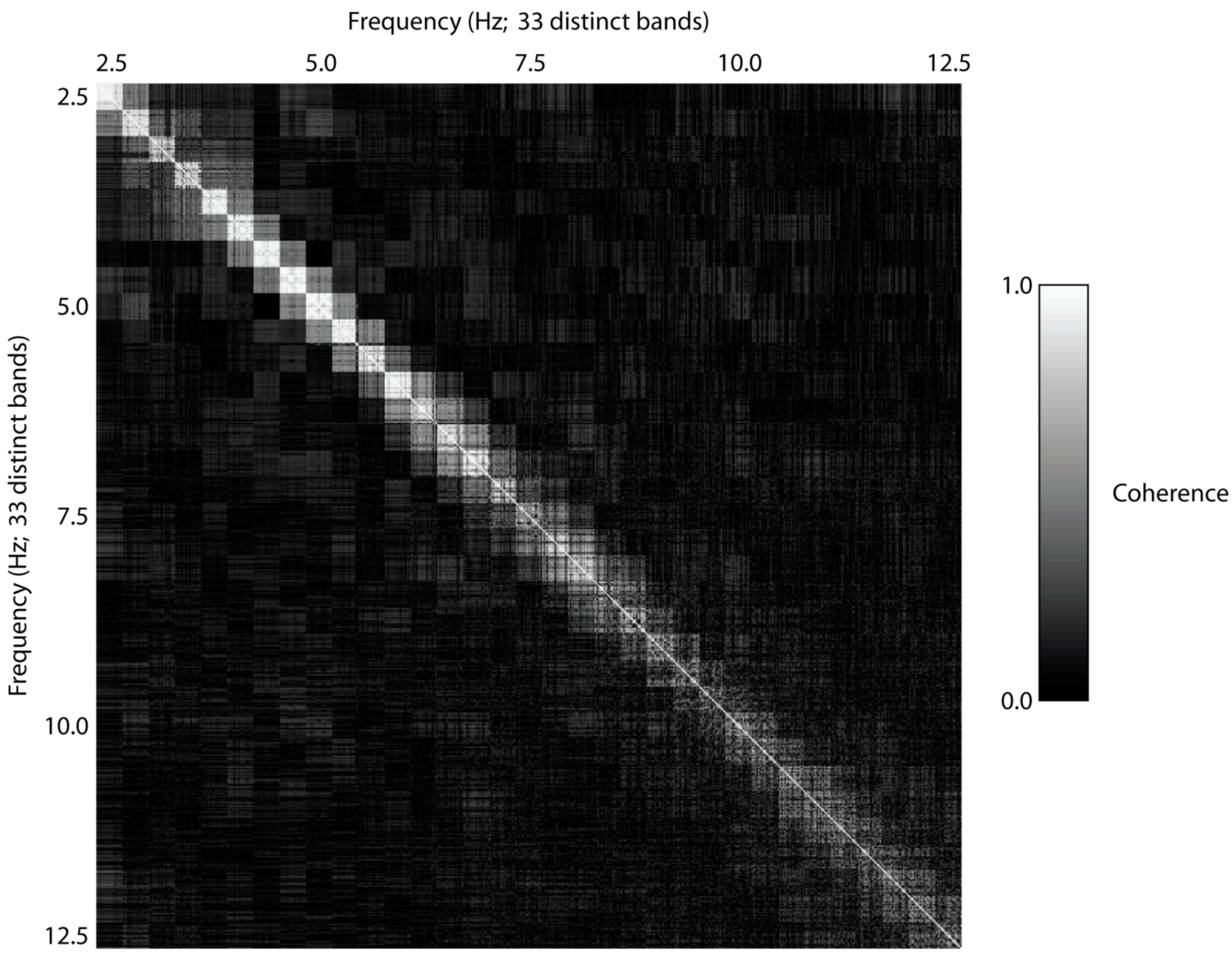


Figure 5

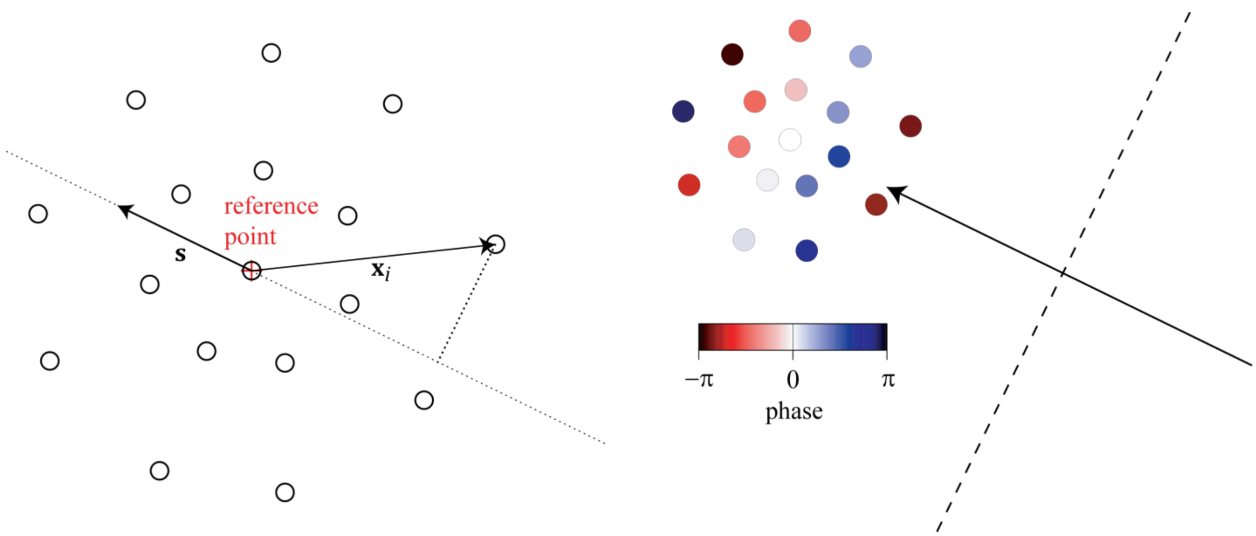
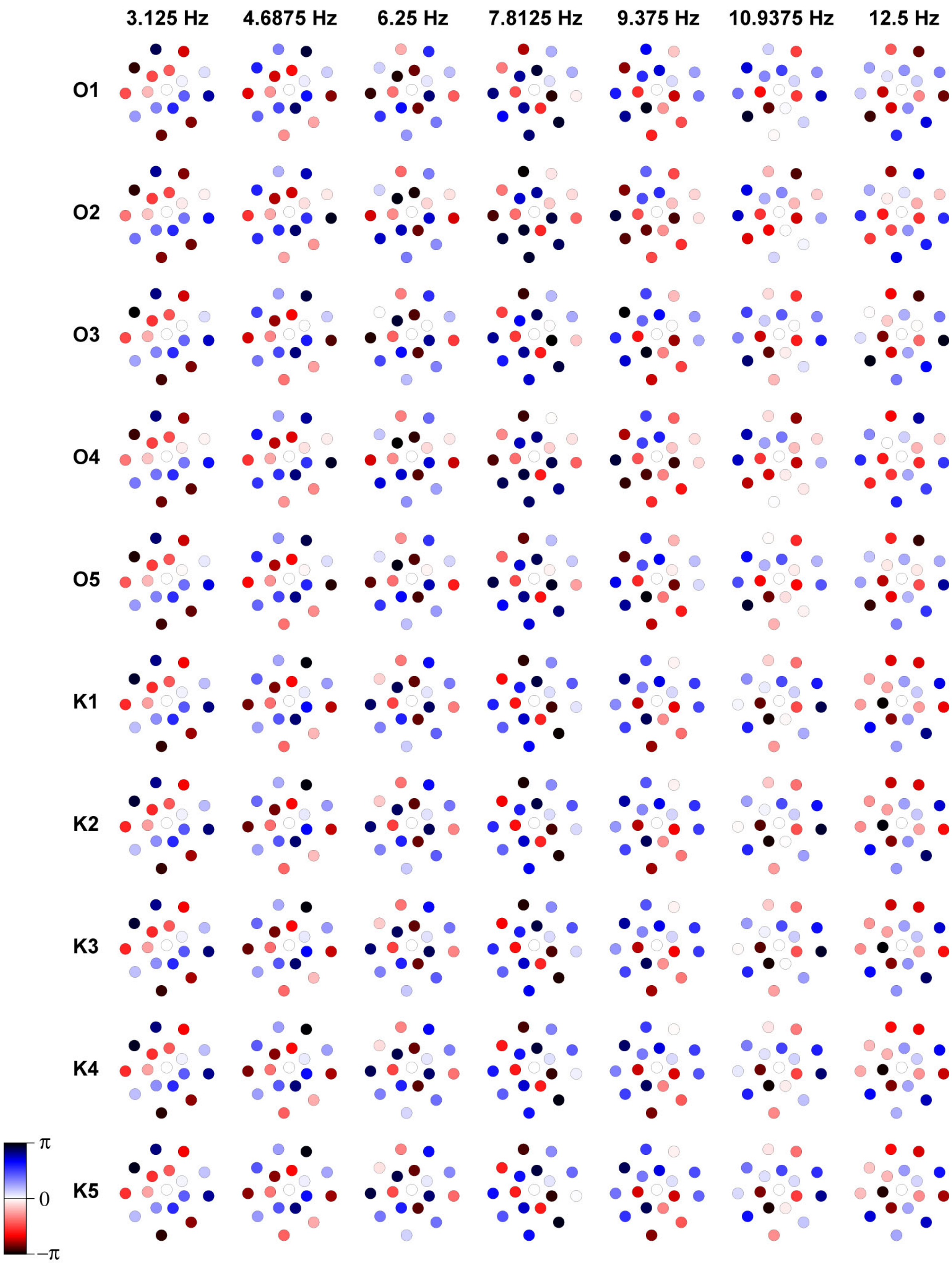
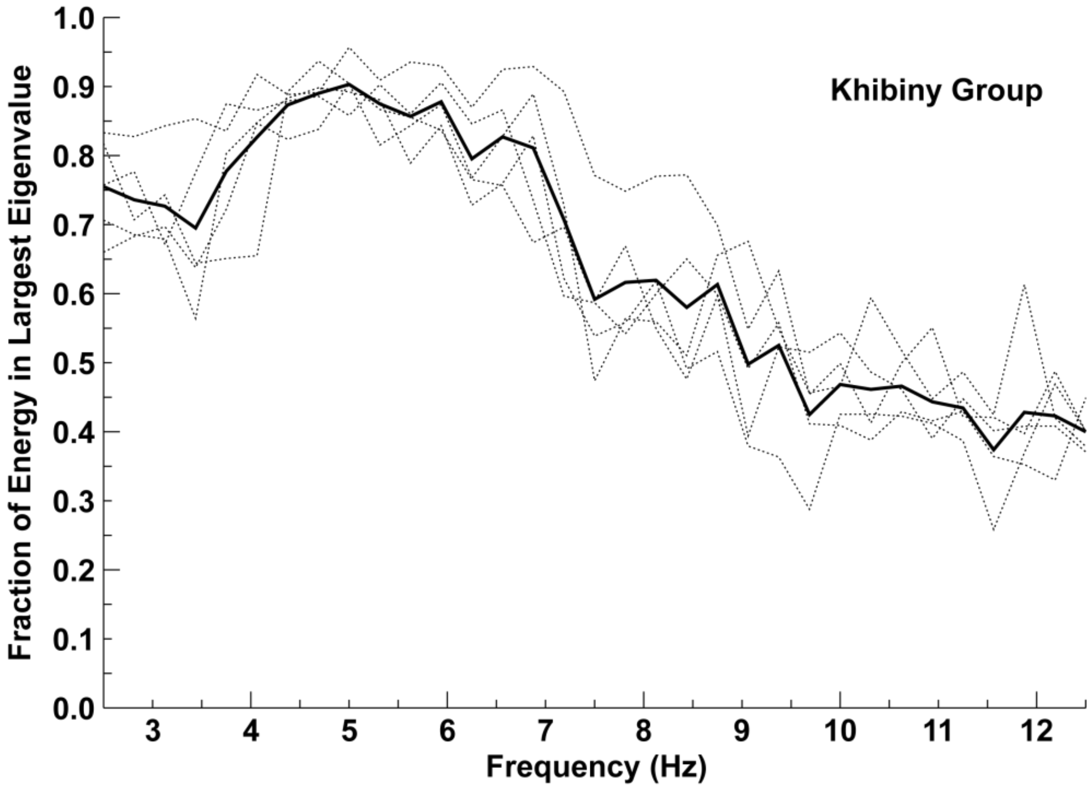


Figure 6



3.125 Hz
4.6875 Hz
6.25 Hz
7.8125 Hz
9.375 Hz
10.9375 Hz
12.5 Hz
O1
O2
O3
O4
O5
K1
K2
K3
K4
K5
π
0
−π

Figure 7


Khibiny Group
Fraction of Energy in Largest Eigenvalue
Frequency (Hz)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
3
4
5
6
7
8
9
10
11
12

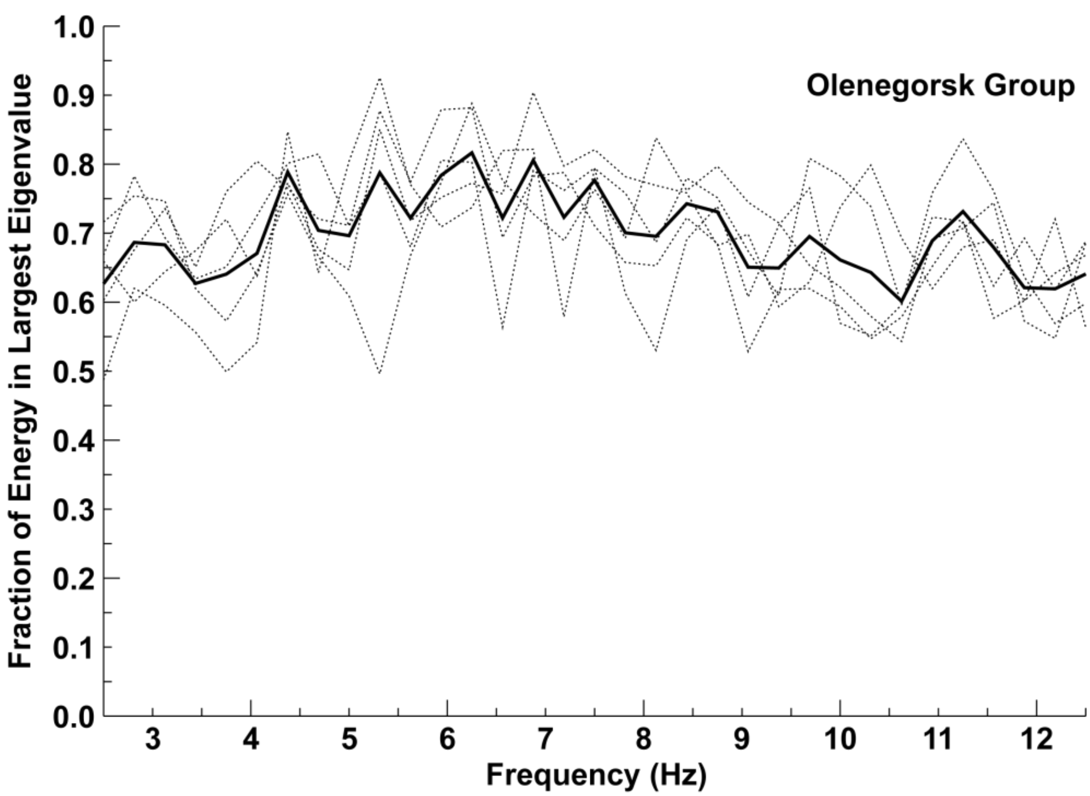

Olenegorsk Group
Fraction of Energy in Largest Eigenvalue
Frequency (Hz)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
3
4
5
6
7
8
9
10
11
12

Figure 8

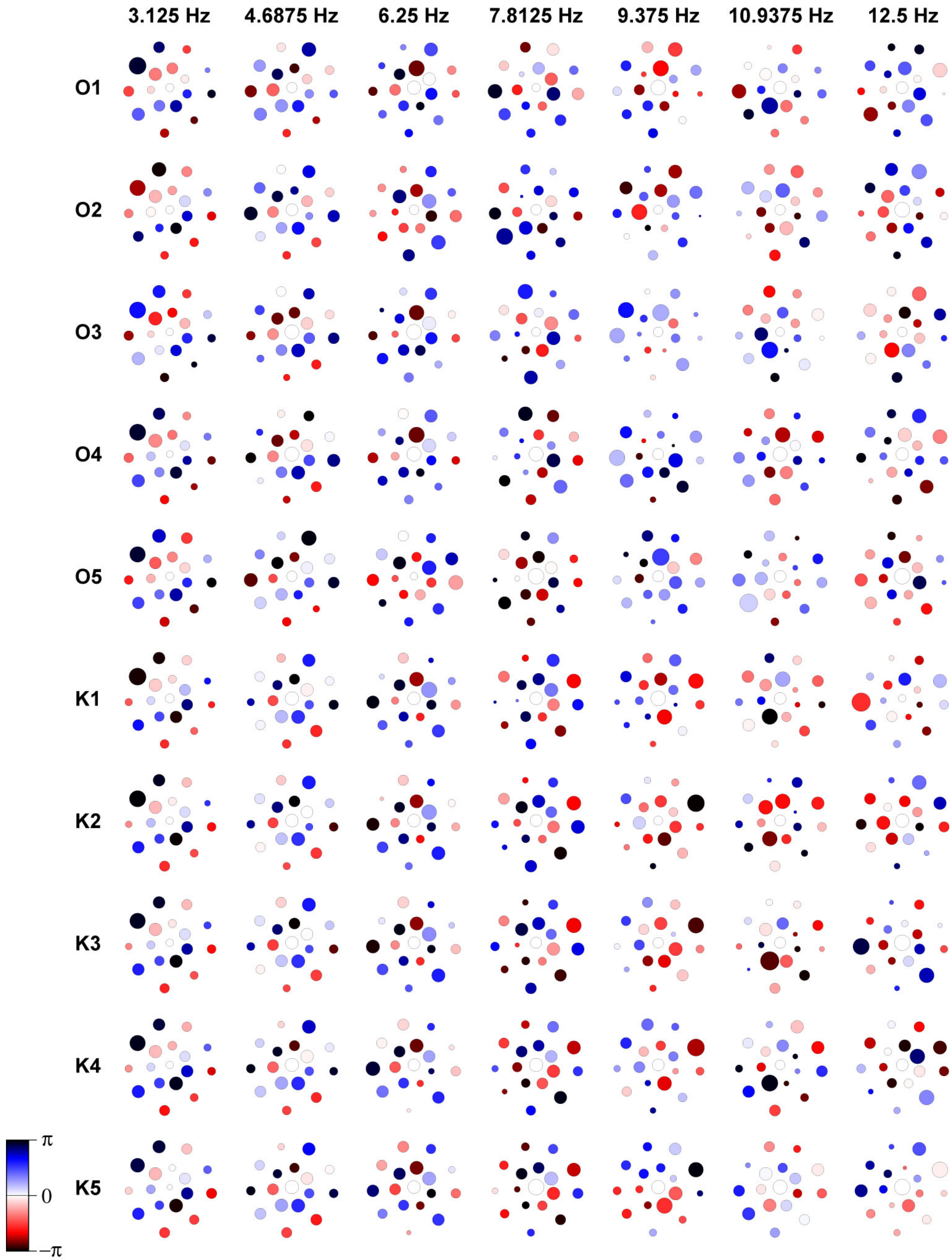
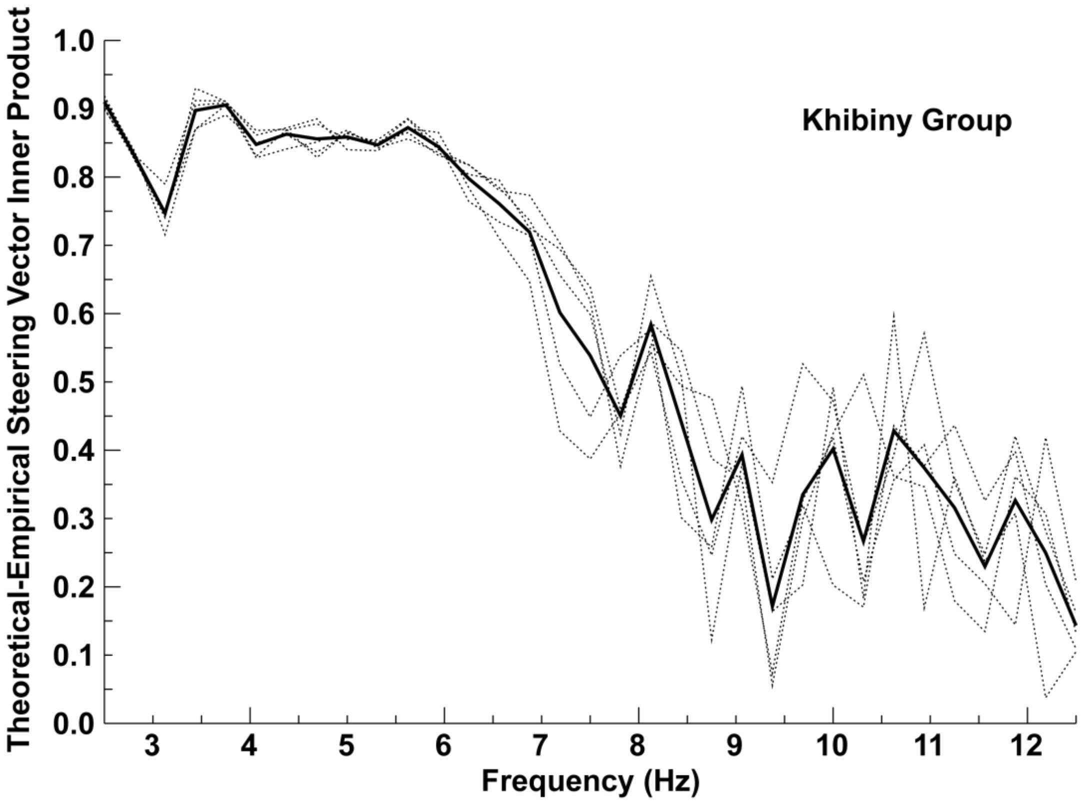


Figure 9


Theoretical-Empirical Steering Vector Inner Product
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
3
4
5
6
7
8
9
10
11
12
Frequency (Hz)
Khibiny Group

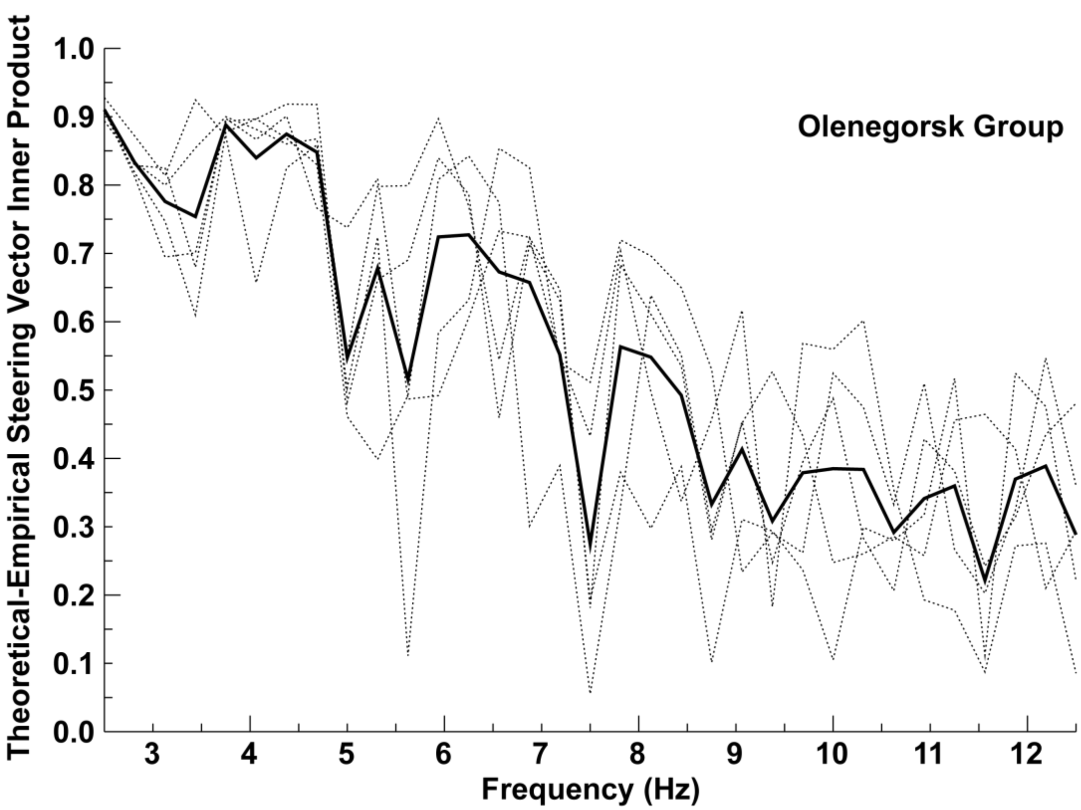

Theoretical-Empirical Steering Vector Inner Product
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
3
4
5
6
7
8
9
10
11
12
Frequency (Hz)
Olenegorsk Group

Figure 10

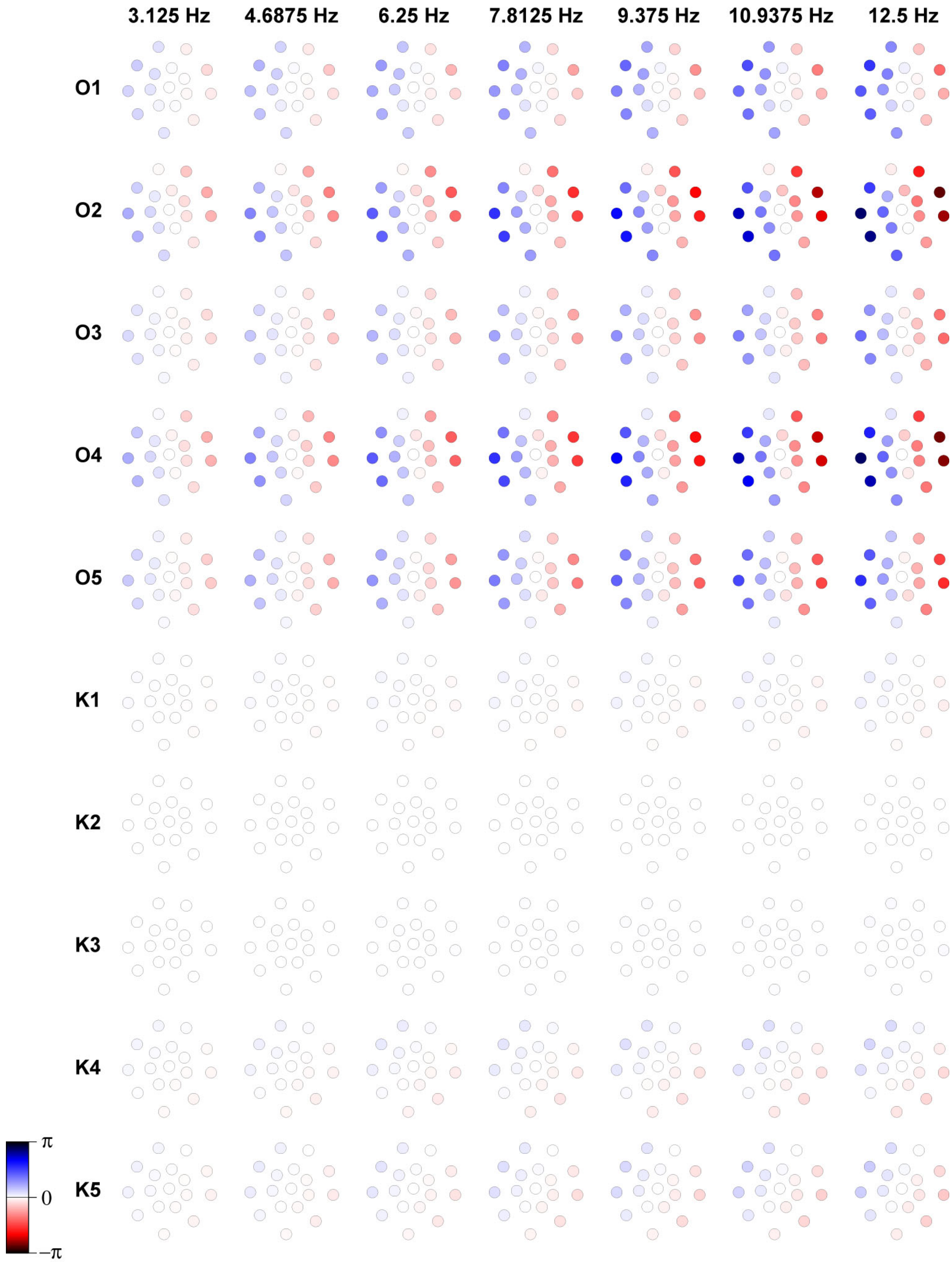
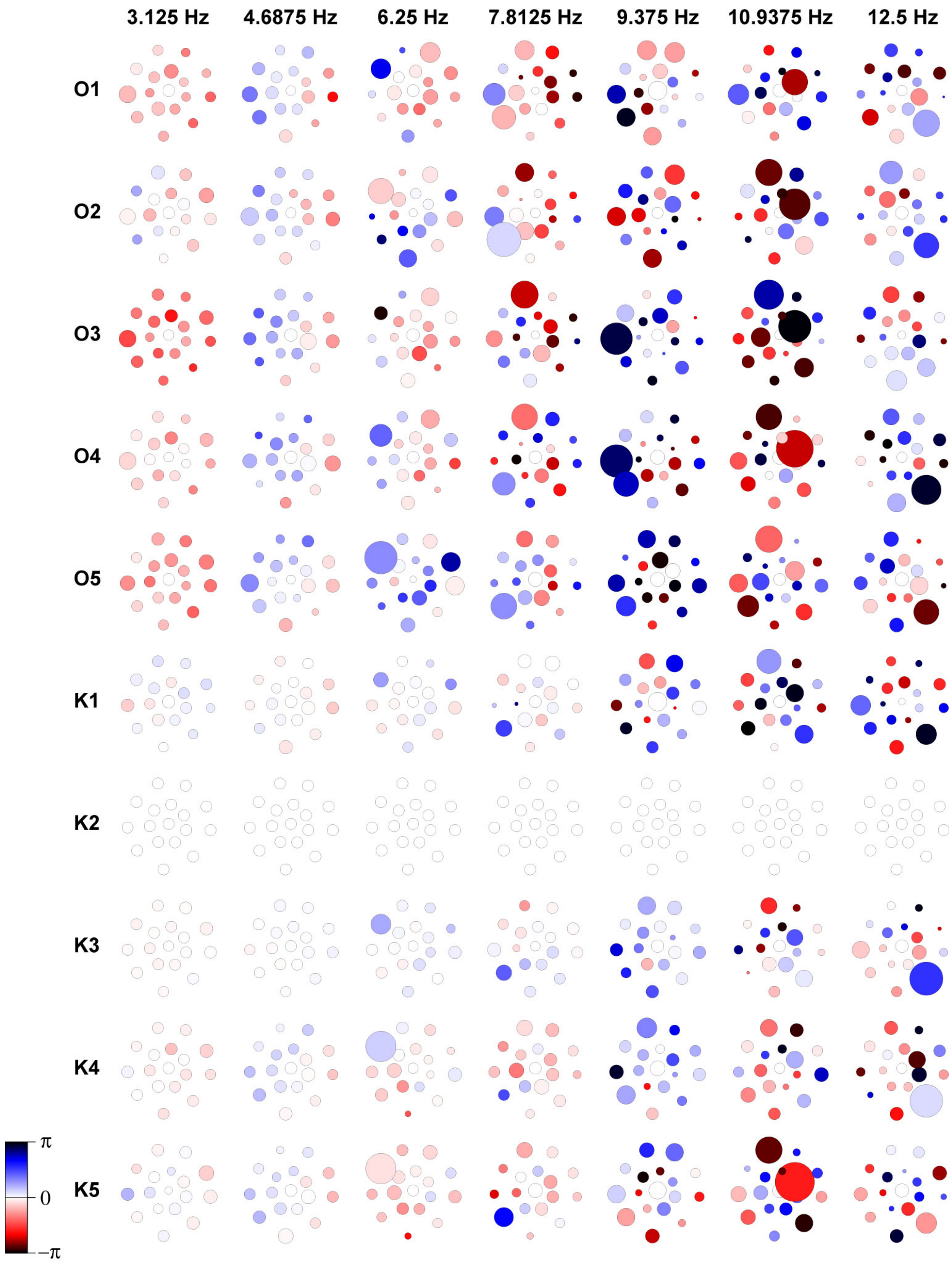

3.125 Hz
4.6875 Hz
6.25 Hz
7.8125 Hz
9.375 Hz
10.9375 Hz
12.5 Hz
O1
O2
O3
O4
O5
K1
K2
K3
K4
K5
π
0
$-\pi$

Figure 11



3.125 Hz
4.6875 Hz
6.25 Hz
7.8125 Hz
9.375 Hz
10.9375 Hz
12.5 Hz
O1
O2
O3
O4
O5
K1
K2
K3
K4
K5
π
0
−π

Figure 12

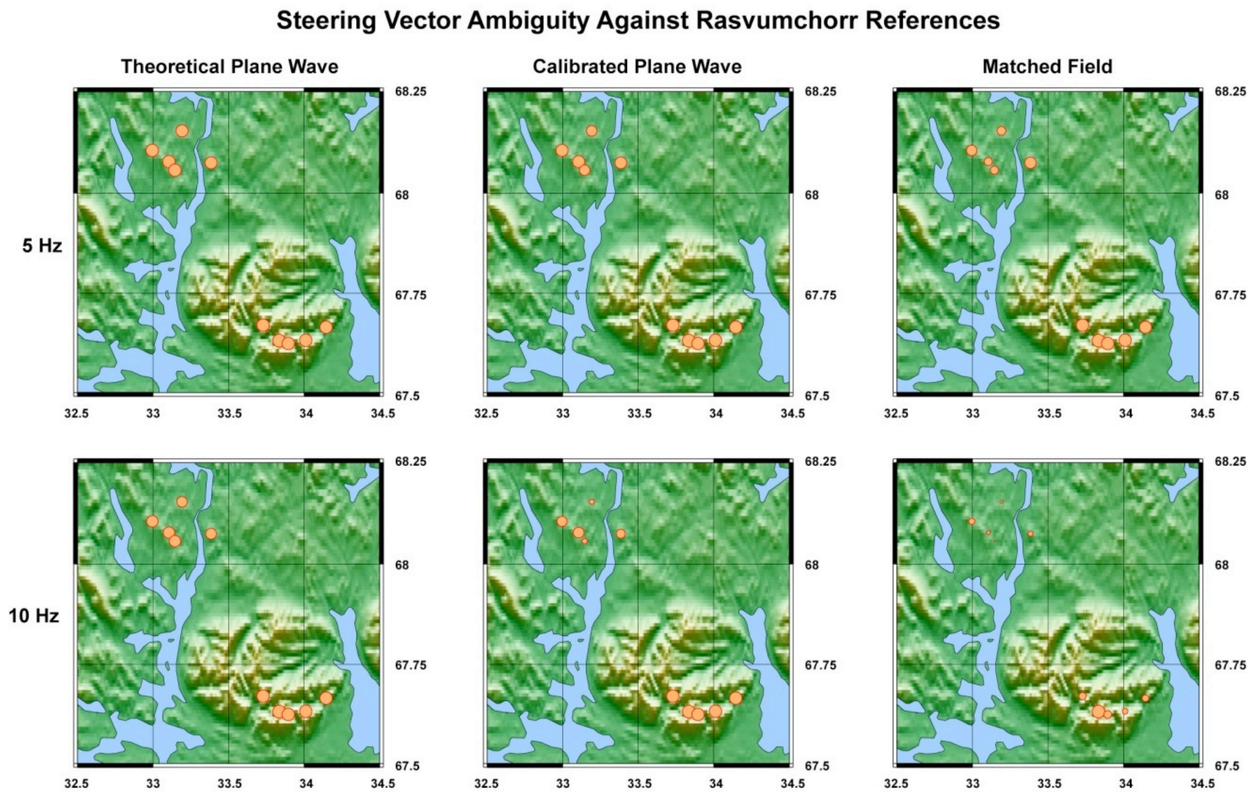
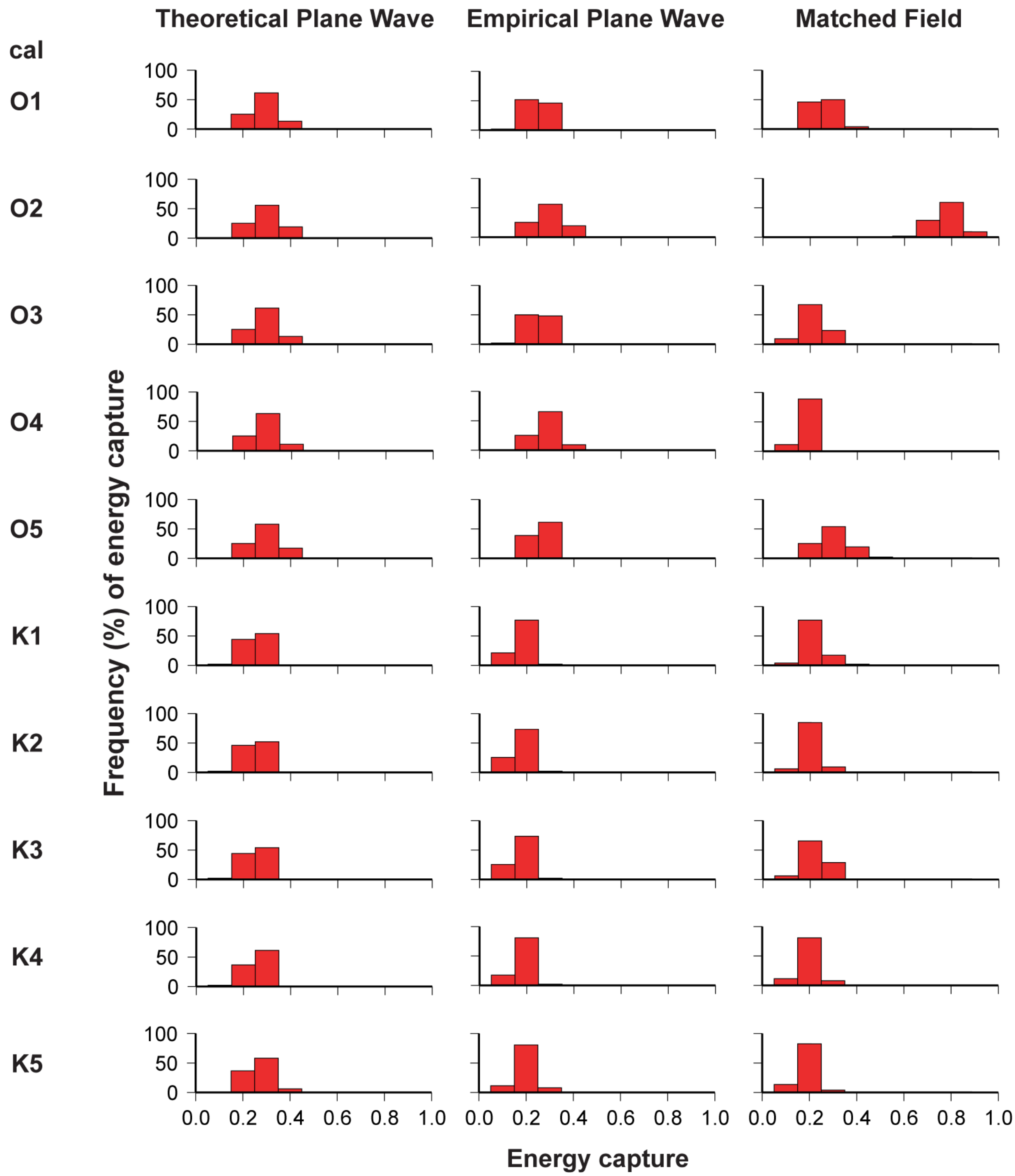


Figure 13



Theoretical Plane Wave
Empirical Plane Wave
Matched Field
cal
O1
O2
O3
O4
O5
K1
K2
K3
K4
K5
Frequency (%) of energy capture
Energy capture
0.0 0.2 0.4 0.6 0.8 1.0
100
50
0

Figure 14

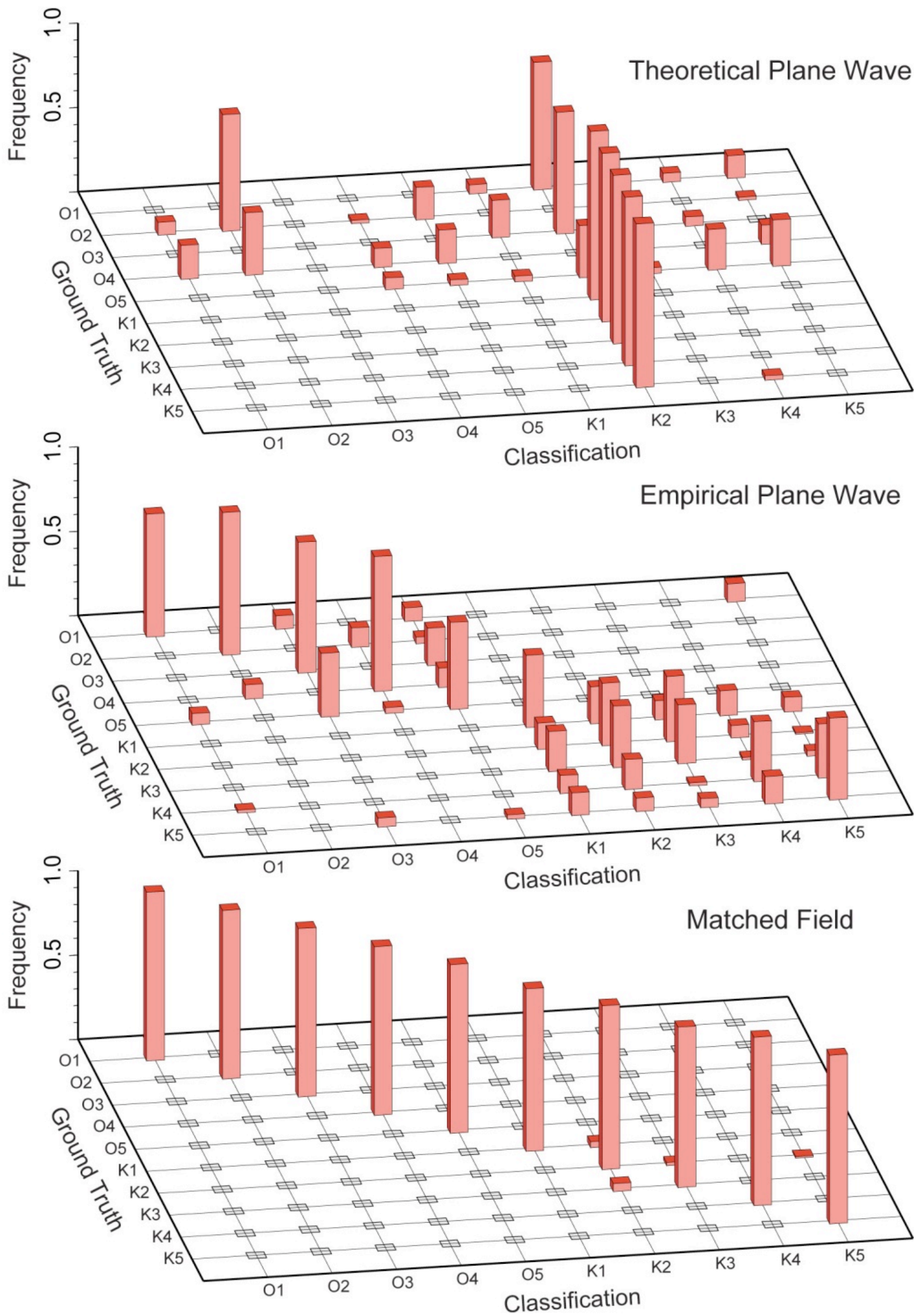

Theoretical Plane Wave
Empirical Plane Wave
Matched Field
Frequency
1.0
0.5
Ground Truth
O1
O2
O3
O4
O5
K1
K2
K3
K4
K5
Classification

Figure 15